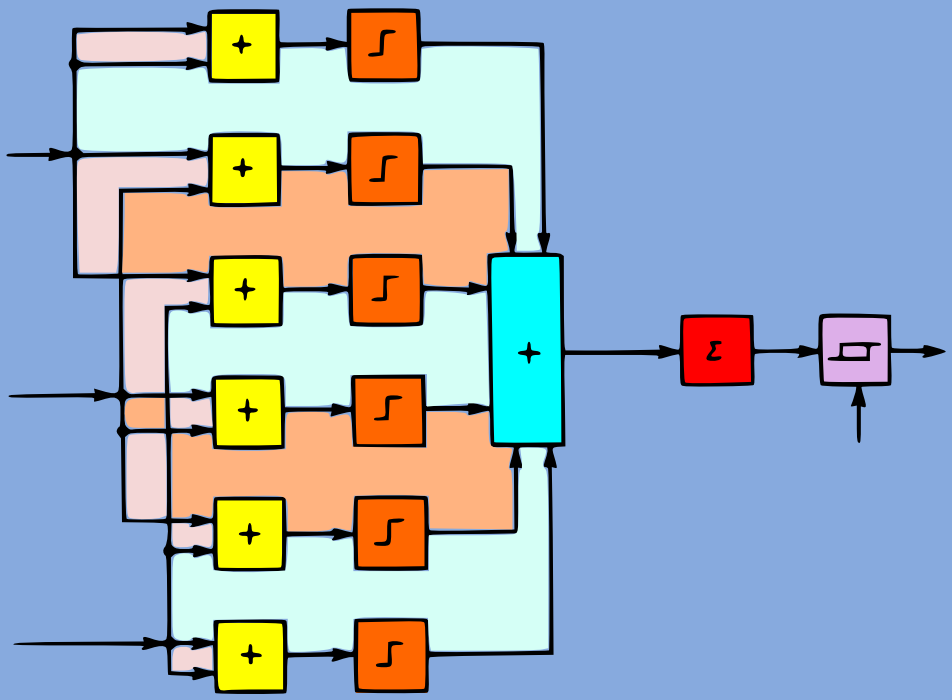
B. Lévine

# Fondements Théoriques de la RADIOTECHNIQUE STATISTIQUE

Tome3



Éditions Mir Moscou

Б. ЛЕВИН

# ТЕОРЕТИЧЕСКИЕ ОСНОВЫ СТАТИСТИЧЕСКОЙ РАДИОТЕХНИКИ

## Том III

СОВЕТСКОЕ РАДИО
МОСКВА

B. LÉVINE

# FONDEMENTS THÉORIQUES DE LA RADIOTECHNIQUE STATISTIQUE

*Tome III*

ÉDITIONS MIR · MOSCOU

Traduit du russe par
A. SOKOVA

*На французском языке*

© Издательство «Советское радио», 1976 г.
© Traduction française Editions Mir 1979

# PRÉFACE À L'ÉDITION FRANÇAISE

Le problème de l'indétermination a priori nous semble assez intéressant pour être traité dans le cadre des problèmes de la synthèse statistique des systèmes optimaux de réception et de traitement de l'information.

En effet, essentiel des points de vue philosophique, méthodologique et pratique de la théorie des probabilités et de la mathématique statistique, ce problème a acquis de nos jours une importance toute particulière car les conceptions probabilistes et statistiques se trouvent matérialisées dans de nombreux domaines techniques.

Nous ne prétendons pas cerner complètement le problème. Cette entreprise aurait abouti à un ouvrage très volumineux et donc ayant peu de chances d'être lu. Nous avons choisi à notre goût certaines approches du problème de la synthèse des systèmes informationnels dans les conditions d'indétermination a priori.

D'abord la méthode adaptative. Certes, elle mérite une attention plus grande qu'il n'a été possible de lui prêter dans cet ouvrage. La seule excuse à ce manque d'attention est qu'il existe une littérature abondante consacrée à cette question. En revanche, nous traitons en détail le principe spécial de l'optimalité asymptotique qui dans de nombreux cas peut avantageusement se substituer ou être associé à la méthode adaptative.

On trouvera également les méthodes traditionnelles qui ont été longtemps utilisées pour surmonter l'indétermination a priori, à savoir les méthodes non paramétriques. Tout récemment encore, ces méthodes semblaient peu attractives pour les applications radiotechniques et dans les télécommunications, mais aujourd'hui, combinées à d'autres principes, elles ouvrent des perspectives intéressantes. Enfin, dans certains cas la théorie classique des décisions offre des procédures efficaces pour surmonter l'indétermination a priori ; c'est le cas par exemple des observations caractérisées par une famille exponentielle de distributions des probabilités.

Un grand nombre de résultats exposés dans ce dernier volume de notre monographie, notamment ceux du troisième chapitre et certains autres sont originaux, ils ont été obtenus par l'auteur et ses

élèves. Certains des résultats mentionnés ont été publiés dans des revues soviétiques et étrangères depuis 1967.

Les matières exposées dans cet ouvrage sont essentiellement basées sur le cours spécial consacré au problème de l'indétermination a priori dans la radiotechnique statistique, professé par l'auteur depuis 1972 à l'Institut Electrotechnique des communications. Cette expérience pédagogique nous a beaucoup aidé dans la préparation de cet ouvrage. Un cours analogue a été donné en 1972 à l'Université technique de Budapest où nous avons eu des discussions fort utiles avec le professeur Sandor C s i b i. Tous ceux qui enseignent savent que ce n'est qu'après avoir traité une question devant un auditoire attentif qu'on acquiert la certitude de la connaître à fond. Cette certitude alimente l'inspiration, si nécessaire pour le travail littéraire.

Dans l'ouvrage sont conservées les désignations et la terminologie adoptées dans les deux premiers volumes parus en français en 1973 dans les Editions « Mir ».

L'auteur serait reconnaissant à tous ceux qui trouveraient possible d'adresser aux Editions « Mir » leurs remarques et suggestions sur l'édition française des trois volumes des « Fondements théoriques de la radiotechnique statistique ».

Lektori benevolo salutem!

*B. LÉVINE*

*Professeur à l'Institut Electrotechnique des communications de Moscou*
*Moscou, janvier* 1979

# INTRODUCTION

Durant les dernières décennies la liaison étroite entre le progrès technique dans le domaine de l'électronique et des méthodes d'étude des systèmes complexes d'un côté, et la théorie mathématique des processus aléatoires et la théorie des décisions de l'autre devient toujours plus évidente. Les succès atteints dans la résolution de certains problèmes importants sont dus à l'utilisation des méthodes probabilistes; le progrès technique a de son côté stimulé le développement des méthodes mentionnées.

A l'heure actuelle on utilise une approche statistique unifiée à de nombreuses recherches scientifiques et aux travaux pratiques concernant les systèmes informationnels de télécommunications terrestres et cosmiques, le radar à grande et petite distance, la séismologie et la télémétrie, la radio-astronomie et le sonar, la commande automatique des systèmes complexes, la météorologie, l'océanographie et le diagnostic médical.

Une fois que le partage des domaines d'application des modèles probabilistes et déterministes est devenu à peu près définitif, on peut procéder à une étude critique de la méthodologie et de certains résultats de la *théorie statistique des systèmes informationnels.*

La formation classique du problème de la *synthèse statistique* du dispositif optimal de réception dans un système de communications, inspirée par les travaux de V. Kotelnikov [8] et basée sur le critère de Bayes, fait appel en outre au critère de qualité et à un grand nombre de caractéristiques a priori des signaux, des bruits et à leur mode d'interaction. Cependant souvent l'ingénieur-chercheur ou bien ne dispose pas de ces données, ou bien il n'est pas sûr qu'elles soient suffisamment vérifiées, ce qui est un obstacle à l'utilisation des résultats de la théorie. De plus, même lorsque l'on surmonte l'indétermination a priori au moyen d'hypothèses subjectives, de critères de qualité simplifiés ou d'une restriction de la classe des modèles utilisés, il est parfois difficile de trouver une solution analytique simple du problème et d'interpréter la solution sous forme de schémas radiotechniques réalisables. En outre, une décision, optimale pour certaines hypothèses (par exemple, pour des bruits normale-

ment répartis), peut cesser de l'être pour d'autres conditions et, par conséquent, pour beaucoup de situations réelles, on a besoin d'un grand nombre de systèmes optimaux, ce qui présente un inconvénient du point de vue économique et pratique.

Au début des années 50, pendant la période « romantique » du développement de la théorie statistique des télécommunications, Woodward notait avec raison que l'obstacle fondamental empêchant l'élaboration d'une théorie satisfaisante de la réception, applicable à des systèmes d'observation tels que le radar est, sans aucun doute, la question des distributions a priori, tant pour les communications, que pour les paramètres non essentiels [3]. La théorie des probabilités et la statistique mathématique sont historiquement nées de la nécessité de résoudre des problèmes dans des conditions d'indétermination ou d'information incomplète, d'analyser les résultats des observations de processus physiques soumis à des *lois statistiques*. Ces observations varient d'un cas à l'autre, de sorte qu'il est impossible de prédire exactement le résultat de chacune des observations prise à part.

Laissant de côté la question philosophique de savoir si les lois de la nature ont un caractère déterministe ou probabiliste, nous voulons souligner certaines particularités méthodologiques des formulations des problèmes pratiques dans les conditions d'indétermination a priori. Parfois, on dispose des distributions de probabilité de toutes les grandeurs entrant dans les conditions du problème données sur l'ensemble de leurs valeurs possibles. On dit que ces problèmes sont à *information a priori complète*. Lorsqu'au contraire les distributions de certaines grandeurs ne sont pas données, on parle de problèmes *à indétermination a priori*.

La théorie de Bayes de vérification des hypothèses statistiques et d'estimation des paramètres inconnus (étudiée dans le tome II du présent ouvrage) n'est pas applicable à la solution des problèmes d'information a priori complète. Mais alors, quelle est l'importance pratique de la théorie de Bayes? Ces questions sont étudiées dans un grand nombre d'ouvrages. Nous allons donner ici deux points de vue: l'un d'entre eux est celui du bien-fondé de l'application de la théorie axiomatique des probabilités à la solution des problèmes de la pratique courante [6, 7, 13], selon l'autre, les distributions probabilistes sont subjectives et ne caractérisent que le *degré de certitude* du chercheur que l'information a priori dont il dispose est fiable [6, 15]. Ce dernier point de vue est très répandu parmi les statisticiens et les chercheurs. Laplace a remarqué que la théorie des probabilités c'est le bon sens incarné par le calcul. Par cette remarque l'éminent déterministe a certainement voulu souligner le rôle décisif que jouent les appréciations subjectives a priori. On a même proposé le terme de théorie *subjective* des décisions statistiques comme synonyme de la théorie de Bayes [5]. Il existe des problè-

mes où l'un des points de vue mentionnés suffit à justifier l'application pratique de la théorie de Bayes.

Cependant, comme nous l'avons noté ci-dessus, il est souvent difficile de donner les répartitions a priori dans les problèmes de la synthèse statistique des systèmes informationnels, et même si l'on surmonte les difficultés a priori au moyen des hypothèses subjectives, on n'arrive pas au résultat désiré. Ainsi, on a vu naître et se développer la théorie de détection, de discrimination, de classification des signaux et d'estimation des paramètres de ces derniers en présence de bruits, qui consiste à élaborer les méthodes permettant de surmonter l'indétermination a priori, d'obtenir des structures voisines des structures optimales, invariantes au changement des modèles des signaux et des bruits et qui soient facilement interprétables dans le langage des éléments radiotechniques.

Dans le présent ouvrage on expose certaines méthodes permettant de surmonter l'indétermination a priori, applicables aux problèmes de détection et de classification des signaux noyés dans des bruits, les bruits étant caractérisés par une large classe de distributions et de modes d'interaction avec le signal.

Dans le premier chapitre on formule les principaux problèmes de la synthèse, on examine les données a priori et on fournit les résultats fondamentaux formulés précédemment dans la théorie de la synthèse statistique avec information a priori complète. Puis on montre comment est surmontée l'indétermination a priori, on cite les *critères de qualité* des algorithmes de détection des signaux dans les conditions d'indétermination paramétrique et non paramétrique a priori, les algorithmes adaptatifs, les algorithmes asymptotiquement optimaux.

Le second chapitre est consacré aux *méthodes non paramétriques*. Après un bref aperçu des méthodes de vérification des hypothèses sur la forme de la fonction de répartition, on donne un exposé systématique de la théorie des algorithmes de signe, de rang, de signe-rang de détection des signaux noyés dans des bruits.

Dans le troisième chapitre on donne un exposé complet de la théorie des détecteurs *asymptotiquement optimaux*. On envisage des échantillons indépendants et corrélés des algorithmes de rang et des algorithmes numériques.

Dans les deux premiers paragraphes du chapitre 4 on traite des algorithmes de classification des observations à partir des *échantillons d'apprentissage* et l'on analyse les propriétés asymptotiques des statistiques de classification. Puis on donne différentes procédures d'apprentissage de Bayes (algorithme de Robbins, algorithmes récurrents utilisant l'approximation stochastique et les méthodes des fonctions potentielles, algorithme adaptatif de détection d'un signal en présence de paramètres perturbateurs, etc.). Le chapitre se termine par la théorie des algorithmes adaptatifs asymptotique-

ment optimaux, fondée sur les résultats exposés dans le chapitre 3.

La théorie des algorithmes asymptotiquement optimaux de détection des signaux fait pour la première fois l'objet d'une monographie. Il est possible que le lecteur remarque une certaine partialité de l'auteur à l'égard de cette théorie. Bien que l'ouvrage fasse une place suffisante aux autres méthodes, la préférence souvent donnée aux algorithmes asymptotiquement optimaux ne s'explique pas seulement par l'apport de l'auteur et de ses élèves à la théorie de ces algorithmes.

Le défaut majeur des algorithmes non paramétriques (de signe, de rang) est justement qu'on ne peut les utiliser que pour des échantillons indépendants et que le choix de tel ou tel algorithme s'inspire de considérations euristiques. La théorie des algorithmes asymptotiquement optimaux permet de comprendre en quel sens l'algorithme non paramétrique donné est optimal. Cette théorie est applicable aussi bien aux échantillons indépendants qu'aux échantillons corrélés. Et bien que les algorithmes asymptotiquement optimaux ne soient pas, en toute rigueur, non paramétriques, il y a une séparation fonctionnelle bien nette de l'influence de la distribution des bruits et de la forme du signal et, lorsqu'on y ajoute un dispositif d'adaptation, ils deviennent non paramétriques.

On sait que l'utilisation des méthodes asymptotiques en statistique permet de trouver, pour une vaste classe de phénomènes, des lois générales ne dépendant pas des propriétés individuelles des représentants de cette classe. En règle générale, ces lois asymptotiques sont des variantes de la loi des grands nombres ou du théorème de la limite centrale de la théorie des probabilités.

La théorie des algorithmes asymptotiquement optimaux de détection des signaux noyés dans des bruits permet de donner un sens nouveau aux résultats bien connus de la théorie de détection de Bayes, obtenus pour le modèle d'un bruit normal additif. De plus, on peut utiliser ces résultats non seulement dans le cas où le modèle du bruit correspond à la situation réelle, mais également pour la synthèse des algorithmes asymptotiquement optimaux de détection des signaux pour des bruits quelconques, car pour la réalisation de ces algorithmes on utilise, dans la plupart des cas, des convertisseurs non linéaires inertiels des statistiques suffisantes dont la distribution asymptotique des probabilités est normale.

Cette dernière conclusion permet de présenter les détecteurs asymptotiquement optimaux sous la forme de deux dispositifs. Dans le premier, on forme les statistiques asymptotiquement suffisantes à partir des données a priori ou à l'aide des échantillons d'apprentissage. Le second se compose d'éléments utilisés dans les détecteurs modernes: filtres adaptés, corrélomètres, intégrateurs, éléments à seuil. La caractéristique du convertisseur non linéaire d'entrée dépend de la distribution du bruit, et si l'on ne connaît

pas cette distribution, on peut obtenir une estimation de la caractéristique mentionnée à l'aide d'un échantillon d'apprentissage.

L'utilisation des algorithmes asymptotiquement optimaux pour des échantillons de taille peu importante (ce qui correspond aux conditions réelles de leur utilisation) est également justifiée, à condition qu'ils convergent vers l'algorithme optimal suffisamment vite. Les algorithmes ainsi obtenus sont souvent même meilleurs que les algorithmes basés sur d'autres principes. Il importe de savoir la taille des échantillons (ou le temps d'observation) pour laquelle la perte en qualité entraînée par l'emploi d'un algorithme asymptotiquement optimal ne dépasse pas une certaine grandeur. On peut résoudre ce problème ardu soit sous forme analytique, soit en accumulant les résultats d'une simulation sur ordinateur. Nous ne toucherons pas à cette question dans notre ouvrage. Nous nous bornerons à dire que nous disposons de données expérimentales portant à croire que dans certains cas la perte relative n'est pas supérieure à quelques pour cent lorsque la taille de l'échantillon est de l'ordre de cinq cents. On peut noter à cet effet un travail récent [14], où l'on étudie le rôle des ordinateurs en tant qu'outil d'expérimentation permettant non pas de démontrer mais de formuler une hypothèse acceptable.

On peut espérer que les algorithmes asymptotiquement optimaux trouveront de vastes domaines d'application. Nous allons seulement mentionner les applications les plus évidentes de notre point de vue, à savoir: les télécommunications cosmiques, la radio-astronomie, les mesures des paramètres des cibles s'éloignant. L'extension du domaine d'application des algorithmes asymptotiquement optimaux exigera, dans certains cas, une augmentation notable de la vitesse de fonctionnement des dispositifs techniques utilisés, afin qu'il soit possible, durant un temps suffisamment court, d'accumuler un échantillon de taille suffisante et d'avoir le temps d'effectuer toutes les opérations prescrites par l'algorithme.

Il y a lieu de souligner enfin que dans le présent ouvrage on se limite aux classes d'algorithmes convergeant « bien » vers les algorithmes optimaux lorsque la taille de l'échantillon observé (ou d'apprentissage) augmente indéfiniment. Nous ne voudrions pas que cette limitation soit comprise comme un principe orthodoxe, excluant toute méthode euristique. Ainsi, par exemple, bien que les algorithmes asymptotiquement optimaux soient une approximation rationnelle des algorithmes optimaux, il est possible qu'on trouve, pour des échantillons de taille finie, une autre méthode d'approche, tout comme l'approximation par des sommes finies de séries divergentes se trouve être parfois meilleure que celle des séries convergentes pour un même nombre de termes de la série. Il en est de même pour le cas de l'utilisation d'estimations non consistantes des paramètres pour des tailles réduites des échantillons ou pour l'affirmation qu'étant donné le petit coefficient d'efficacité asymptotique relative

l'utilisation d'un algorithme non paramétrique n'est pas efficace pour un échantillon de petite taille.

Les problèmes de la synthèse optimale des algorithmes de détection et de classification des signaux pour un intervalle de temps d'observation limité (échantillon de taille finie) sont étudiés dans le chapitre 5. Pour les résoudre, on fait appel aux principes de similitude et d'invariance, parfaitement efficaces dans les conditions d'indétermination paramétrique en présence de bruits additifs normaux corrélés, de puissance moyenne inconnue ou de spectre énergétique à paramètres inconnus.

Dans le présent ouvrage, seuls sont étudiés en détail les problèmes de la détection des signaux noyés dans des bruits (problème de la vérification d'une hypothèse contre une alternative). Cependant, les méthodes proposées pour surmonter l'indétermination a priori peuvent également être utilisées dans des problèmes plus généraux de discrimination des signaux (vérification des hypothèses à alternatives multiples), ainsi que dans les problèmes de détection simultanée et d'estimation des paramètres. Dans la bibliographie des chapitres correspondants, on peut trouver les références d'ouvrages où figurent des généralisations détaillées.

Toute monographie a certaines limites: son volume, le délai d'édition, le public, les intérêts scientifiques de l'auteur, etc. C'est pourquoi de nombreux problèmes pratiques intéressants touchant au choix de la décision dans des conditions d'indétermination a priori se sont trouvés en dehors du cadre de cet ouvrage. Il y a lieu avant tout de citer à cet égard le problème de prise d'une décision dans des situations de conflit, lorsque l'information a priori inconnue n'est pas soumise à l'effet des circonstances objectives mais subit l'influence d'un adversaire. Ce problème est envisagé comme un « jeu » avec un adversaire (la nature) dont, dans le cas général, on ne connaît pas la stratégie. Parfois la stratégie du chercheur est fondée sur l'hypothèse que l'adversaire choisit toujours la répartition la moins favorable des paramètres du problème. Cette stratégie conduit à la règle du minimax de choix d'une décision.

La théorie des décisions de Bayes peut être considérée comme une partie spéciale de la théorie des jeux lorsque le « jeu » est mené contre un adversaire « faible » dont la stratégie, dans le sens probabiliste, est connue du chercheur. Parmi l'abondante littérature consacrée à la théorie des jeux, on peut recommander [2] et pour une étude plus approfondie [4] où le lecteur trouvera des références supplémentaires.

Les problèmes que nous envisageons se rapportent aux systèmes informationnels ouverts qui sont une idéalisation admissible pour un grand nombre de problèmes du radar et des télécommunications. Par contre, les systèmes de commande automatique sont le plus souvent des systèmes fermés que l'on ne peut pas ramener à des

systèmes ouverts. L'insuffisance d'information a priori conduit à la nécessité d'étudier simultanément le phénomène à commander et la commande. Les actions de commande qui dans ces situations revêtent un caractère double portent le nom de commande duale. Les questions de la commande duale, basées sur la théorie des décisions statistiques, sont étudiées dans la monographie fondamentale [11] (voir également [12]). Aux problèmes de la commande optimale doivent leur développement les méthodes mathématiques nouvelles de résolution de problèmes variationnels, notamment le principe du maximum de Pontriaguine et la méthode de programmation dynamique [1, 9, 10].

La théorie de la synthèse des systèmes informationnels en cas d'indétermination est loin d'être achevée. Les succès les plus importants étaient toujours liés, et le seront dans l'avenir, à notre avis, aux méthodes statistiques, bien qu'on doive être prêt à faire face aux difficultés. Ainsi par exemple, on note dans [16] que des difficultés de ce genre apparaissent lors de l'analyse de systèmes complexes dont un élément est l'homme ou une collectivité, aussi propose-t-on dans cet article une nouvelle approche du problème du choix d'une décision, inspirée de la théorie des ensembles non distincts.

A l'affirmation pessimiste de M. Zadeh « plus notre analyse du problème est profonde et plus sa solution devient incertaine » on peut opposer les admirables paroles du poète : « Plus la nuit est noire et plus les étoiles sont brillantes ».

CHAPITRE PREMIER

# PROBLÈME DE L'ÉLIMINATION DE L'INDÉTERMINATION A PRIORI

## 1.1. PROBLÈMES DE LA SYNTHÈSE AVEC INFORMATION A PRIORI COMPLÈTE

**1.1.1. Terminologie.** Nous allons tout d'abord introduire la terminologie dont nous nous servirons dans tout l'exposé ultérieur. Il est très important de donner une définition précise des termes utilisés, nous en avons vu déjà une partie dans les premières pages du présent tome.

Un *système informationnel* est un moyen technique de description, de traitement, de transmission, de réception, de conservation et de perception des communications, c'est-à-dire d'une multitude de renseignements utilisés par l'homme pour connaître l'Univers et lui-même. Un système informationnel peut être une partie d'un système plus important, mais le critère de qualité de son fonctionnement doit être donné d'une manière autonome. Un système informationnel fonctionne conformément à un algorithme imposé par ce critère et comprenant le choix d'une décision *). Un certain nombre de systèmes informationnels particuliers ont été mentionnés dans l'introduction.

La *synthèse* consiste à déterminer l'algorithme de fonctionnement d'un système informationnel suivant un critère de qualité donné et à réaliser cet algorithme à l'aide de dispositifs techniques.

Lors de l'*analyse* il faut calculer les caractéristiques techniques des systèmes informationnels.

Les informations arrivant dans un système informationnel sont des *processus aléatoires*. Une communication que l'on connaît entièrement à l'avance (déterministe) n'a pas de valeur informationnelle. En radiotechnique, dans les télécommunications, la commande automatique et autres domaines techniques les messages sont transmis par des *signaux* d'origine physique différente (électrique, électromagnétique, optique, acoustique, etc.). Les signaux sont accompagnés de *bruits*. Tant les signaux que les bruits sont en général des processus aléatoires soumis à des lois de distribution des probabilités.

---

*) Au lieu du terme « algorithme » on emploie parfois l'expression « règle de choix d'une décision »; en statistique mathématique on utilise les termes « test », « critère ».

Nous appellerons les grandeurs (fonctions), obtenues à partir des lois de distribution données des signaux et des bruits, caractéristiques *probabilistes* (a priori). Les estimations (décisions), obtenues sur la base d'échantillons de taille finie ou de réalisations de processus observés sur un intervalle de temps fini sont appelées caractéristiques *statistiques* (a posteriori).

Lorsque l'on utilise des modèles probabilistes de signaux et de bruits on se trouve toujours en présence d'une indétermination. Cependant nous parlerons de problèmes de synthèse statistique à *information a priori complète*, lorsqu'on connaît toutes les fonctions de répartition des probabilités des signaux et des bruits ainsi que certaines limitations supplémentaires impliquées par la formulation du problème. Au contraire, s'il manque certaines données probabilistes, nous parlerons de problèmes de synthèse statistique dans des conditions de l'*indétermination a priori*. Remarquons que si l'on ne dispose d'aucune donnée a priori, la synthèse optimale du système informationnel est impossible. Cependant, en réalité, le chercheur dispose toujours d'un certain nombre de données (même très limité) suggérées par son expérience et son intuition.

**1.1.2. Principaux types de problèmes de la synthèse statistique.** Détection d'un signal noyé dans un bruit. Le problème est de déterminer l'algorithme de traitement du processus observé $x(t)$ qui ne peut être que le bruit $\zeta(t)$ ou une combinaison du signal $s(t)$ et du bruit $\zeta(t)$. Cet algorithme permet de décider sur la présence ou l'absence du signal dans le processus observé. Ce problème se formule en termes de vérification de l'hypothèse statistique (simple ou composite) $H_0$: $x(t) = \zeta(t)$ contre l'alternative (qui peut elle aussi être simple ou composite) $H_1$: $x(t) = \zeta(t) \otimes s(t)$ (le symbole $\otimes$ désigne une interaction quelconque entre le bruit et le signal).

Les caractéristiques de travail de l'algorithme de détection sont les relations entre le risque moyen, la probabilité de fausse alarme et de perte du signal et les données initiales du problème.

Discrimination des signaux noyés dans le bruit. Le problème consiste à trouver l'algorithme de traitement du processus observé $x(t)$ qui est une combinaison du bruit et d'un signal $s_k(t)$ appartenant à un ensemble de signaux $s_0(t), \ldots, s_m(t)$. L'algorithme doit permettre de dire lequel des signaux est présent dans le processus observé. Ce problème est une variante à alternatives multiples du problème de vérification des hypothèses statistiques $H_k$: $x(t) = \zeta(t) \otimes s_k(t)$, $k = 0, \ldots, m$.

Les caractéristiques de l'algorithme de discrimination sont le risque moyen, ainsi que la probabilité de « confusion » des signaux.

Séparation des signaux noyés dans le bruit. Le problème consiste à trouver l'algorithme de traitement du processus observé qui est une combinaison du signal et du bruit,

tel qu'il donne l'estimation d'une ou de plusieurs caractéristiques inconnues du signal. Ces caractéristiques peuvent être des nombres ou des fonctions du temps, de la fréquence, des grandeurs ou des processus aléatoires. Les caractéristiques de l'algorithme de séparation sont la variance, l'écart et la covariance des estimations, le risque moyen.

Détection simultanée (discrimination) des signaux et estimation de leurs paramètres en présence de bruits. Il y a lieu de déterminer l'algorithme de traitement du processus observé, qui est une combinaison du signal (ou d'un signal appartenant à un ensemble de signaux) et du bruit ou seulement le bruit, appelé à répondre aux deux questions suivantes fonctionnellement liées: si, oui ou non, le signal (un signal appartenant à un certain ensemble) est présent dans le processus observé et quelle est la valeur de la caractéristique du signal (estimation des paramètres). Dans le problème envisagé l'estimation doit être réalisée dans les conditions d'incertitude quant à la présence du signal, contrairement au problème où les paramètres du signal sont estimés dans l'hypothèse que celui-ci est présent dans le processus observé.

**1.1.3. Critères de qualité et données a priori nécessaires (détection et discrimination des signaux)** *). Dans les problèmes de la discrimination (détection) des signaux on utilise le critère du *minimum du risque moyen* (critère de Bayes). La règle de Bayes optimale de discrimination de $m+1$ signaux, c'est-à-dire de la partition optimale de l'espace euclidien échantillionné X à $n$ dimensions en $m+1$ domaines disjoints $X_k$, $k=0, \ldots, m$, est basée sur la minimisation du risque moyen

$$R=\sum_{j=0}^{m} p_j r_j=\sum_{j=0}^{m}\sum_{k=0}^{m} p_j \Pi_{jk} \int_{X_k} w_n(\mathbf{x}|s_j)\,d\mathbf{x},$$

$$\bigcup_{k=0}^{m} X_k = X, \tag{1.1}$$

où $(\Pi_{jk})$ est la matrice des pertes; $p_j$ la probabilité a priori de la présence du signal $s_j$, $j=0, \ldots, m$; $w_n(\mathbf{x} \mid s_j)$ la densité de probabilité conditionnelle (fonction de vraisemblance) de l'échantillon $\mathbf{x}=(x_1, \ldots, x_n)$ dans l'hypothèse de la présence du signal $s_j$, et $r_j$ le risque conditionnel.

Ainsi, pour l'utilisation du critère d'optimum de Bayes un grand nombre de données a priori sont nécessaires, à savoir, la

*) Dans les pp. 1.1.3 à 1.1.12 on trouvera un bref aperçu des résultats fondamentaux de la théorie de la synthèse statistique dans les conditions d'information a priori complète (voir, par exemple, le tome II ainsi que [1, 6, 10]).

matrice des pertes, les probabilités a priori de présence des signaux, les modèles des signaux et des bruits déterminant les fonctions de vraisemblance de l'échantillon. Soulignons que l'hypothèse $H_k$ sur la présence du signal $s_k$ est également une information a priori.

Lorsque les probabilités a priori $p_j$, $j = 0, \ldots, m$, ne sont pas données on utilise le *critère du minimax*. L'algorithme du minimax est celui dont la valeur des maxima du risque conditionnel $r_j$, $j = 0, \ldots, m$, est minimale par rapport à d'autres algorithmes. L'algorithme du minimax est un cas particulier de l'algorithme de Bayes pour la distribution $p_j^*$, $j = 0, \ldots, m$, pour laquelle le risque de Bayes

$$R(p_0^*, \ldots, p_m^*) \geqslant R(p_0, \ldots, p_m).$$

Lorsque l'on ne connaît pas la matrice des pertes $(\Pi_{ij})$ on utilise le *critère du maximum de la probabilité a posteriori*; il s'agit de choisir la plus grande des valeurs des probabilités a posteriori des signaux pour l'échantillon donné $\mathbf{x}$

$$P\{s_j|\mathbf{x}\} = \frac{p_j w_n(\mathbf{x} \mid s_j)}{\sum\limits_{k=0}^{m} p_k w_n(\mathbf{x} \mid s_k)}, \qquad j = 0, \ldots, m. \tag{1.2}$$

Si l'on ne connaît ni les probabilités $p_j$, ni la matrice des pertes, on utilise le *critère du maximum de vraisemblance* qui consiste à choisir la plus grande des valeurs de la fonction de vraisemblance $w_n(\mathbf{x} \mid s_j)$, $j = 0, \ldots, m$.

Il y a lieu de souligner que la synthèse des algorithmes optimaux de discrimination suivant l'un quelconque des critères mentionnés n'est possible que si l'on dispose de données a priori permettant de calculer la fonction de vraisemblance $w_n(\mathbf{x} \mid s_j)$, $j = 0, \ldots, m$.

Dans les problèmes de détection (cas binaire: hypothèse $H_0$ — il n'y a pas de signal, hypothèse $H_1$ — le signal est présent) on utilise le *critère de Neyman-Pearson*. La procédure consiste à se donner la probabilité de fausse alarme

$$\alpha = \int\limits_{\mathbf{X}_1} w_n(\mathbf{x}|H_0)\,d\mathbf{x} \tag{1.3}$$

et à minimiser la probabilité de perte du signal

$$\beta = \int\limits_{\mathbf{X}_0} w_n(\mathbf{x}|H_1)\,d\mathbf{x}. \tag{1.3'}$$

Pour utiliser le critè e de Neyman-Pearson il n'est pas nécessaire de connaître la probabilité a priori de présence du signal ni la grandeur des pertes dues aux décisions erronées. Mais les modèles des signaux et des bruits (fonctions de répartition $w_n(\mathbf{x} \mid H_i)$, $i = 0; 1$) doivent être complètement déterminés (dans le sens probabiliste) et, de plus, on doit connaître la probabilité des fausses alarmes.

**1.1.4. Critères de qualité et données a priori indispensables (estimation des paramètres).** Le critère du minimum du risque moyen (critère de Bayes) est également utilisé pour l'estimation des paramètres non observés du signal d'après l'échantillon observé $\mathbf{x} = (x_1, \ldots, x_n)$ du mélange du signal et du bruit. Soient $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_m)$ un vecteur des paramètres aléatoires d'un signal à densité de probabilité mutuelle $w_m(\boldsymbol{\vartheta})$ donnée, et $L_{\mathbf{x}}(\boldsymbol{\vartheta}) = w_n(x_1, \ldots, x_n \mid \boldsymbol{\vartheta})$ la fonction de vraisemblance de l'échantillon. L'estimation de Bayes optimale $\hat{\boldsymbol{\vartheta}}$ du paramètre vectoriel $\boldsymbol{\vartheta} \in \Theta$ s'obtient par minimisation du risque moyen

$$R = \int_{\Theta} \int_{\mathbf{X}} w_m(\boldsymbol{\vartheta}) \Pi(\hat{\boldsymbol{\vartheta}}, \boldsymbol{\vartheta}) L_{\mathbf{x}}(\boldsymbol{\vartheta}) \, d\mathbf{x} \, d\boldsymbol{\vartheta}, \tag{1.4}$$

où $\Pi(\hat{\boldsymbol{\vartheta}}, \boldsymbol{\vartheta})$ est la fonction de pertes. En vertu de (1.4), pour trouver les estimations de Bayes des paramètres des signaux il faut connaître la distribution a priori des paramètres, la fonction de pertes et la fonction de vraisemblance de l'échantillon, cette dernière étant définie à partir des modèles connus du signal et des bruits.

Minimiser le risque moyen $R$ revient à minimiser le *risque a posteriori*

$$J = \int_{\Theta} \Pi(\hat{\boldsymbol{\vartheta}}, \boldsymbol{\vartheta}) W(\boldsymbol{\vartheta} \mid \mathbf{x}) \, d\boldsymbol{\vartheta}. \tag{1.4'}$$

Lorsque la fonction de pertes n'est pas donnée, pour obtenir les estimations on peut utiliser le *critère du maximum de densité a posteriori* du paramètre vectoriel du signal *)

$$W(\boldsymbol{\vartheta} \mid \mathbf{x}) = \frac{w_m(\boldsymbol{\vartheta}) L_{\mathbf{x}}(\boldsymbol{\vartheta})}{\int_{\Theta} w_m(\boldsymbol{\vartheta}) L_{\mathbf{x}}(\boldsymbol{\vartheta}) \, d\boldsymbol{\vartheta}}. \tag{1.5}$$

Pour les estimations d'après une réalisation de $x(t)$ observée sur un intervalle de temps fini, la fonction de vraisemblance dans (1.4) et (1.5) est remplacée par la fonctionnelle du rapport de vraisemblance $l[x(t) \mid \boldsymbol{\vartheta}]$.

Lorsque l'on ne connaît pas la distribution a priori $w_m(\boldsymbol{\vartheta})$, on peut utiliser la méthode du minimax. On recherche une estimation telle que la borne supérieure exacte des valeurs de la fonction du risque conditionnel

$$r(\boldsymbol{\vartheta}) = \int_{\mathbf{X}} \Pi(\hat{\boldsymbol{\vartheta}}, \boldsymbol{\vartheta}) L_{\mathbf{x}}(\boldsymbol{\vartheta}) \, d\mathbf{x} \tag{1.6}$$

ne dépasse pas les bornes supérieures des valeurs de $r(\boldsymbol{\vartheta})$ quelles que soient les estimations.

*) On sait que les estimations suivant le critère du maximum de probabilité a posteriori sont des cas particuliers des estimations de Bayes pour des hypothèses spéciales (voir tome II, page 122).

Parfois on suppose que les paramètres du signal sont des constantes inconnues. Lors de la recherche des estimations de ces paramètres on utilise le *critère du maximum de vraisemblance* en vertu duquel l'estimation la meilleure $\hat{\vartheta}$ du paramètre vectoriel $\vartheta$ est celle pour laquelle est atteinte la borne supérieure exacte des valeurs de la fonction de vraisemblance $L_x(\vartheta)$.

Dans les problèmes de filtrage (d'estimation) d'un signal aléatoire non observable $\xi_s(t)$ d'après une réalisation de $x(t)$ du mélange du signal et du bruit observée sur un certain intervalle de temps on utilise le *critère du minimum du carré moyen de l'erreur* $m_1\{[\hat{\xi}_s(t) - \xi_s(t)]^2\}$, où $\hat{\xi}_s(t)$ est l'estimation du signal qui est une fonctionnelle de la réalisation observée. Dans certains cas ce critère coïncide avec celui du maximum de densité de probabilité a posteriori du signal $\xi_s(t)$ après l'observation d'une réalisation de $x(t)$.

Dans les problèmes de détection simultanée et dans ceux d'estimation des paramètres on utilise également le *critère de Bayes.*

**1.1.5. Types d'algorithmes.** Les solutions des problèmes de la détection, de la discrimination et de la séparation des signaux se présentent sous forme d'*algorithmes* (règles) de transformation des observations (échantillons, réalisations) donnant lieu à des prises de décision ou de formation d'estimations. Autrement dit, les solutions et les estimations sont basées sur des statistiques, c'est-à-dire des fonctions des données échantillonnées ou des fonctionnelles de réalisations. Les caractéristiques probabilistes de ces statistiques sont des caractéristiques de travail des algorithmes correspondants.

Un algorithme est dit *consistant* s'il donne une probabilité d'erreur tendant vers zéro lorsque la taille de l'échantillon augmente indéfiniment. Un algorithme de détection, optimal suivant le critère de Neyman-Pearson, est consistant si, pour une probabilité de fausse alarme donnée, la probabilité de détection correcte tend vers l'unité lorsque la taille de l'échantillon augmente indéfiniment. La *consistance d'une estimation* signifie sa convergence en probabilité vers le paramètre estimé.

Il y a lieu de distinguer le traitement *discret* et *analogique* des observations. Dans le premier cas les observations sont présentées sous forme d'échantillons de taille finie, c'est-à-dire par les valeurs de la réalisation observée du processus aléatoire à des instants donnés *). Dans le second cas les observations sont représentées par la réalisation entière du processus aléatoire sur un intervalle de temps donné.

*) Dans le traitement discret on distingue le traitement analogo-discret où l'on procède uniquement à une quantification temporelle, et le traitement numérique où les observations sont quantifiées en temps et en amplitude.

Les algorithmes peuvent être *simples*, lorsque la décision est prise pour des tailles données de l'échantillon ou un intervalle donné d'observation, et à plusieurs étapes, lorsque l'on peut remettre la prise de décision jusqu'à réception de données supplémentaires. Dans ce dernier cas la taille de l'échantillon ou le temps d'observation sont aléatoires et l'on prend pour critère de qualité de la détection du signal le minimum des tailles moyennes de l'échantillon (tant pour l'hypothèse que pour l'alternative) avant la prise de décision pour une qualité donnée de l'algorithme.

Les algorithmes *récurrents* n'utilisant que les dernières observations et une certaine statistique de toutes les observations antérieures sont d'un grand intérêt pratique. Ces algorithmes permettent de simplifier notablement la réalisation des dispositifs de détection et d'estimation.

Parfois on utilise des *algorithmes d'apprentissage* auxiliaires servant à former les estimations des caractéristiques inconnues des phénomènes. Souvent on se limite d'avance à une classe donnée d'algorithmes, par exemple à des *algorithmes linéaires* qui sont des fonctions linéaires des valeurs échantillonnées ou des fonctionnelles linéaires des réalisations observées.

Dans le cas des processus haute fréquence, la vitesse de fonctionnement limitée des moyens techniques dont on dispose peut être un obstacle à la réalisation des algorithmes optimaux. On procède alors à une détection préalable du processus observé, c'est-à-dire que l'on sépare l'enveloppe ou la phase à variation lente d'un processus à bande étroite, puis on procède à un *traitement optimal postdétecteur*.

**1.1.6. Fonction de décision.** Il est parfois commode de formuler l'algorithme de détection à l'aide de la *fonction de décision* $\Phi(\mathbf{x})$ donnée dans l'espace des échantillons $\mathbf{X}$:

$$\Phi(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \in \mathbf{X}_1, \\ 0, & \mathbf{x} \in \mathbf{X}_0, \end{cases} \tag{1.7}$$

où $\mathbf{X}_1$ est le domaine d'adoption de l'hypothèse $H_1$ correspondant à la présence du signal (domaine critique); $\mathbf{X}_0$ le domaine d'adoption de l'hypothèse $H_0$ — absence de signal.

La fonction de décision (1.7) correspond à l'algorithme *non randomisé*. Pour un algorithme randomisé cette fonction peut prendre une valeur quelconque de zéro à l'unité et indique la probabilité d'adoption de l'hypothèse $H_1$ *).

*) Plus exactement, si l'on observe $\mathbf{x}$, on est en présence d'une expérience aléatoire à deux issues possibles: $A$ et $\bar{A}$, dont les probabilités sont $\Phi(\mathbf{x})$ et $1 - \Phi(\mathbf{x})$. Si $A$ apparaît, on adopte l'hypothèse $H_1$ (voir [5]).

Les probabilités de fausse alarme et de perte de signal s'expriment à l'aide de la fonction de décision comme suit:

$$\alpha = m_1 \{\Phi(\mathbf{x}) | H_0\} = \int_{\mathbf{X}} \Phi(\mathbf{x})\, w_n(\mathbf{x}|H_0)\, d\mathbf{x}, \tag{1.8}$$

$$\beta = 1 - m_1 \{\Phi(\mathbf{x}) | H_1\} = 1 - \int_{\mathbf{X}} \Phi(\mathbf{x})\, w_n(\mathbf{x}|H_1)\, d\mathbf{x}. \tag{1.8'}$$

Pour l'algorithme non randomisé donné par la fonction de décision (1.7), les formules (1.8) et (1.8') coïncident avec les formules (1.3) et (1.3').

En utilisant (1.7) on peut écrire l'expression (1.1) du risque moyen pour le cas binaire de la manière suivante:

$$R = q\Pi_{00} + p\Pi_{10} - \int_{\mathbf{X}} \Phi(\mathbf{x}) [p(\Pi_{10} - \Pi_{11})\, w_n(\mathbf{x}|H_1) - \\ - q(\Pi_{01} - \Pi_{00})\, w_n(\mathbf{x}|H_0)]\, d\mathbf{x}, \tag{1.8''}$$

où $p = 1 - q$ est la probabilité a priori de présence du signal.

**1.1.7. Modèles du signal et du bruit.** On utilise souvent le modèle d'un *signal quasi déterministe du type*

$$s(t) = \sum_{i=1}^{m} \vartheta_i \varphi_i(t), \tag{1.9}$$

où $\vartheta_1, \ldots, \vartheta_m$ est un ensemble de paramètres inconnus (aléatoires) et $\varphi_1(t), \ldots, \varphi_m(t)$ sont des fonctions données. En introduisant le vecteur des paramètres *) $\boldsymbol{\vartheta}' = (\vartheta_1, \ldots, \vartheta_m)$ et le vecteur fonction $\boldsymbol{\varphi}(t) = \begin{pmatrix} \varphi_1(t) \\ \cdots \\ \varphi_m(t) \end{pmatrix}$ on peut écrire le signal $s(t)$ comme le produit du vecteur ligne $\boldsymbol{\vartheta}'$ par le vecteur colonne $\boldsymbol{\varphi}(t)$

$$s(t) = \boldsymbol{\vartheta}' \boldsymbol{\varphi}(t), \tag{1.9'}$$

c'est-à-dire comme une fonction linéaire des paramètres inconnus. Un modèle plus général admet une fonction non linéaire du paramètre $\boldsymbol{\vartheta}$. Des cas particuliers du signal (1.9) sont un signal déterministe $s(t)$, un signal d'amplitude aléatoire $\vartheta s(t)$, un signal harmonique d'amplitude et de phase aléatoires

$$s(t) = a \cos(\omega_0 t - \varphi) = a \cos\varphi \cos\omega_0 t + a \sin\varphi \sin\omega_0 t,$$
$$\vartheta_1 = a\cos\varphi, \quad \vartheta_2 = a \sin\varphi,$$
$$\varphi_1(t) = \cos\omega_0 t, \quad \varphi_2(t) = \sin\omega_0 t. \tag{1.10}$$

---

*) Ici et plus bas l'apostrophe désigne la transposition.

Dans les systèmes multiplex on voit apparaître des signaux $\mathbf{s}(t) = (s_1(t), \ldots, s_N(t))$ vectoriels quasi déterministes qui, généralisant (1.9′) peuvent s'écrire comme suit:

$$\mathbf{s}(t) = \boldsymbol{\vartheta}'\boldsymbol{\Phi}(t), \tag{1.11}$$

où $\boldsymbol{\vartheta}'$ est le vecteur ligne de dimension $1 \times m$ des paramètres aléatoires, et $\boldsymbol{\Phi}(t)$ la matrice de dimension $m \times N$ des fonctions de base des signaux.

On utilise également les modèles des signaux *stochastiques*: le processus markovien multiplement connexe et le processus aléatoire normal. Dans ce dernier cas un signal stochastique peut s'écrire sur l'intervalle $(-T, T)$ comme

$$s(t) = \sum_{j=1}^{\infty} \vartheta_j \frac{\varphi_j(t)}{\sqrt{\lambda_j}}, \qquad |t| \leqslant T, \tag{1.12}$$

où $\lambda_j > 0$, $\varphi_j(t)$ sont les valeurs propres et les fonctions orthonormées propres de l'équation intégrale linéaire

$$\varphi(t) = \lambda \int_{-T}^{T} B(t, y)\, \varphi(y)\, dy, \quad |t| \leqslant T, \tag{1.13}$$

de noyau égal à la fonction de corrélation $B(t, y)$ du processus aléatoire normal $s(t)$, et $\{\vartheta_j\}$ sont des grandeurs aléatoires normales indépendantes égales à

$$\vartheta_j = \sqrt{\lambda_j} \int_{-T}^{T} s(t)\, \varphi_j(t)\, dt. \tag{1.14}$$

Le *bruit blanc* est un modèle très usité du bruit. C'est un processus aléatoire hypothétique à spectre énergétique uniforme et de fonction de corrélation

$$B(t - y) = N_0 \delta(t - y), \tag{1.15}$$

où $N_0$ est la densité spectrale constante du bruit. Le bruit blanc normal (de densité spectrale égale à l'unité) peut être défini comme la dérivée (dans le sens général) d'un processus de Wiener, c'est-à-dire d'un processus aléatoire normal $\xi(t)$ à accroissements indépendants, pour lequel on a

$$m_1\{\xi(t + \tau) - \xi(t)\} = 0, \tag{1.16}$$

$$m_1\{[\xi(t + \tau) - \xi(t)]^2\} = |\tau|. \tag{1.16′}$$

Un *bruit stationnaire normal corrélé* $\xi(t)$ de moyenne nulle est entièrement défini par sa fonction de corrélation $B_\xi(\tau) =$

$= m_1 \{\xi(t)\, \xi(t+\tau)\}$, car sa distribution multidimensionnelle est

$$w_n(\mathbf{x}, \mathbf{t}) = (2\pi)^{-n/2} \sqrt{\det \mathbf{M}^{-1}} \exp\left\{-\frac{1}{2}\mathbf{x}'\mathbf{M}^{-1}\mathbf{x}'\right\}, \qquad (1.17)$$

où $\mathbf{x}' = (x_1, \ldots, x_n)$; $\mathbf{t}' = (t_1, \ldots, t_n)$; $\mathbf{M}^{-1}$ est la matrice inverse de la matrice de corrélation $\mathbf{M}$. Les éléments de la matrice carrée $\mathbf{M}$ de dimension $n \times n$ sont

$$M_{ij} = B_\xi(t_i - t_j), \qquad i, j = 1, \ldots, n.$$

On prend également pour modèle de processus aléatoires et de bruits les *processus markoviens* qui sont des processus aléatoires sans post-action. La caractéristique probabiliste d'un processus markovien est la densité de probabilité de transition $w_2(x_{i+1}, t_{i+1} \mid x_i, t_i)$, car la densité de probabilité d'ordre quelconque de ce processus s'obtient à partir de la *condition de factorisation*

$$w_n(x_1, \ldots, x_n; t_1, \ldots, t_n) = w_1(x_1, t_1) \prod_{i=1}^{n-1} w_2(x_{i+1}, t_{i+1} \mid x_i, t_i), \qquad (1.18)$$

la densité unidimensionnelle $w_1(x, t)$ satisfaisant à la condition de compatibilité (voir tome I, 4.5.1). Un processus normal stationnaire markovien est un processus aléatoire normal de fonction de corrélation exponentielle $B(\tau) = B(0)\, e^{-\lambda|\tau|}$.

Un *processus markovien vectoriel* $\xi(t) = (\xi_1(t), \ldots, \xi_N(t))$ est caractérisé par un ensemble de densité de probabilité de transition $w(\mathbf{x}_{i+1}, \mathbf{t}_{i+1} \mid \mathbf{x}_i, \mathbf{t}_i)$, $i = 1, 2, \ldots$, et la densité de probabilité conjointe de ses composantes est donnée par la formule (1.18) à condition d'y remplacer les arguments scalaires par des arguments vectoriels.

Une généralisation du modèle markovien simple est un *processus markovien de connexité $k$*, c'est-à-dire un processus aléatoire à post-action finie *). La densité de probabilité conditionnelle de transition dépend dans ce cas non pas de un, mais de $k$ valeurs antérieures du processus. En désignant

$$\mathbf{x}_i^j = (x_i, \ldots, x_j), \quad \mathbf{t}_i^j = (t_i, \ldots, t_j), \qquad (1.19)$$

on peut écrire la condition de factorisation de la densité de probabilité multidimensionnelle d'un processus markovien $k$-connexe sous

*) Une suite markovienne $k$-connexe peut être considérée comme une suite markovienne vectorielle

$$\mathbf{x}_j = (x_{j-k}, \ldots, x_{j-1}), \quad j = 1, 2, \ldots$$

la forme

$$w_{n+k}(\mathbf{x}^n_{-k+1},\ \mathbf{t}^n_{-k+1}) = w_k(\mathbf{x}^0_{-k+1},\ \mathbf{t}^0_{-k+1}) \times$$
$$\times \prod_{i=1}^{n} w(x_i,\ t_i | \mathbf{x}^{i-1}_{i-k},\ \mathbf{t}^{i-1}_{i-k}). \tag{1.20}$$

Parfois on introduit dans les conditions des problèmes de détection, de discrimination et de séparation des signaux les paramètres des signaux, des bruits et des effets concomitants, ne présentant pas d'intérêt par eux mêmes, mais influant sur la distribution des statistiques utilisées pour prendre une décision, car la distribution initiale, à laquelle satisfait l'échantillon observé, dépend entre autres des paramètres mentionnés. Ces paramètres sont dits *perturbateurs*, et les problèmes correspondants sont des *problèmes à paramètres perturbateurs*. Ce sont, par exemple, l'amplitude et la phase initiale d'oscillations harmoniques haute fréquence utilisées en radiotechnique en qualité de signaux porteurs d'information utile. La puissance des bruits est un autre exemple de paramètre perturbateur.

Mentionnons enfin deux modèles essentiels d'interaction d'un signal et d'un bruit: le modèle additif et le modèle multiplicatif. Dans le premier cas le bruit s'ajoute au signal, et dans le second il est multiplié par le signal. A titre d'exemple de bruit additif on peut citer le bruit de fluctuation, tandis que l'hétérogénéité du milieu traversé par le signal provoque un bruit multiplicatif.

**1.1.8. Algorithmes optimaux de détection et de discrimination des signaux.** Pour tous les critères de qualité mentionnés, l'algorithme optimal de détection du signal déterministe $s(t)$ noyé dans le bruit $\zeta(t)$ d'après l'échantillon observé $\mathbf{x} = (x_1, \ldots, x_n)$, $x_i = x(t_i)$, consiste à comparer avec un seuil le *rapport de vraisemblance*

$$l(\mathbf{x}) = \frac{W_n(\mathbf{x} \mid H_1)}{W_n(\mathbf{x} \mid H_0)}, \tag{1.21}$$

où $H_0$: $x(t) = \zeta(t)$, $H_1$: $x(t) = \zeta(t) \otimes s(t)$.

Lors de la détection d'après une réalisation $x(t)$ du processus observé dans le cas régulier (voir tome II, 3.4.3) c'est la fonctionnelle du rapport de vraisemblance

$$l[x(t)] = \lim_{n\to\infty} l(x_1, \ldots, x_n),\ t \in (-T,\ T) \tag{1.22}$$

qui est comparée avec un seuil.

Dans le cas d'un signal quasi déterministe $s(t, \vartheta)$ on calcule tout d'abord le rapport de vraisemblance conditionnel $l(\mathbf{x} \mid \vartheta)$ ou la fonctionnelle du rapport de vraisemblance $l[x(t) \mid \vartheta]$, puis

on prend les moyennes sur $\vartheta$

$$\Lambda(\mathbf{x}) = \int_{\Theta} l(\mathbf{x}|\vartheta)\, w_m(\vartheta)\, d\vartheta, \tag{1.23}$$

$$\Lambda[x(t)] = \int_{\Theta} l[x(t)|\vartheta]\, w_m(\vartheta)\, d\vartheta \tag{1.24}$$

que l'on compare avec les seuils.

Dans le problème de distinction des hypothèses $H_k$: $x(t) = \zeta(t) \otimes s(t)$, $k = 0, \ldots, m$, on calcule $m$ rapports de vraisemblance (ou $m$ fonctionnelles du rapport de vraisemblance)

$$l_k(\mathbf{x}) = \frac{W_n(\mathbf{x} \mid H_k)}{W_n(\mathbf{x} \mid H_0)}, \qquad k = 1, \ldots, m,$$

le dispositif à seuil étant remplacé par un dispositif de comparaison, choisissant le rapport de vraisemblance maximal *).

Il y a lieu de souligner que le rapport de vraisemblance (fonctionnelle du rapport de vraisemblance) est une *statistique suffisante*, c'est-à-dire une fonction de l'échantillon observé $x_1, \ldots, x_n$ (de la réalisation $x(t)$) qui permet une *réduction des données* sans perte d'information. Grâce à la réduction l'espace euclidien des échantillons à $n$ dimensions (espace hilbertien des réalisations) du problème de la détection du signal se trouve appliqué sur un ensemble *unidimensionnel* des valeurs d'une grandeur aléatoire (non négative): rapport de vraisemblance (fonctionnelle du rapport de vraisemblance). La décision optimale concernant la présence du signal est prise lorsque cette grandeur aléatoire dépasse un certain seuil dépendant du critère choisi, et sa répartition des probabilités détermine la qualité de la décision prise (probabilité d'erreur).

**1.1.9. Algorithmes optimaux de détection et de discrimination des signaux noyés dans un bruit normal additif.** L'algorithme optimal de détection (de discrimination) des signaux déterministes noyés dans un bruit additif normal corrélé peut être réalisé à l'aide d'un filtre adapté généralisé (ensemble de filtres) et d'un dispositif à seuil (dispositif de comparaison). La valeur du processus à la sortie du filtre adapté en fin d'observation est une fonctionnelle linéaire $\int_{-T} V(t)\, x(t)\, dt$ de la réalisation observée $x(t)$ dont le noyau $V(t)$

*) Un tel système de discrimination des signaux n est optimal que dans un cas particulier. L'algorithme général de Bayes de discrimination des signaux est donné dans le second tome de l'ouvrage.

est donné par l'équation intégrale linéaire

$$\int_{-T}^{T} B_\xi (t-u) V(u)\, du = s(t), \quad |t| \leqslant T, \tag{1.25}$$

et dépend du type du signal $s(t)$ et de la fonction de corrélation du bruit $B_\xi(\tau)$.

La variante discrète de l'algorithme optimal découle de l'expression du logarithme du rapport de vraisemblance

$$\ln l(x) = \mathbf{s}'\mathbf{M}^{-1}\mathbf{x} - \frac{1}{2}\mathbf{s}'\mathbf{M}^{-1}\mathbf{s}, \tag{1.26}$$

où $\mathbf{x}' = (x_1, \ldots, x_n)$; $\mathbf{s}' = (s_1, \ldots, s_n)$; $s_i = s(t_i)$, $t_i \in (-T, T)$ et $\mathbf{M}$ est la matrice de corrélation du bruit normal. En vertu de cet algorithme, il y a lieu de prendre une somme pondérée des valeurs échantillonnées du processus et de comparer la somme obtenue avec un seuil, c'est-à-dire

$$\mathbf{U}'\mathbf{x} \geqslant C + \frac{d^2}{2}. \tag{1.27}$$

où

$$d^2 = \mathbf{s}'\mathbf{M}^{-1}\mathbf{s}. \tag{1.28}$$

Les coefficients de pondération sont égaux aux composantes du produit $\mathbf{U}' = \mathbf{s}'\mathbf{M}^{-1}$ du vecteur des valeurs du signal $\mathbf{s}'$ par la matrice de corrélation inverse $\mathbf{M}^{-1}$ du bruit. La statistique $\mathbf{U}'\mathbf{x} = \mathbf{s}'\mathbf{M}^{-1}\mathbf{x}$ dans (1.27) est une grandeur aléatoire normale de paramètres $(0, d^2)$ en l'absence de signal et $(d^2, d^2)$ lorsque le signal est présent *).

Les probabilités de fausse alarme $\alpha$ et de détection correcte du signal $1-\beta$ se calculent à l'aide de l'intégrale tabulée de Laplace $F(x)$ et dépendent du rapport généralisé signal/bruit $d^2$

$$\left.\begin{matrix}\alpha \\ 1-\beta\end{matrix}\right\} = 1 - F\left(\frac{C}{d} \pm \frac{d}{2}\right), \tag{1.29}$$

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}}\, du. \tag{1.30}$$

L'algorithme de détection (de discrimination) peut être généralisé au cas d'un signal quasi déterministe qui représente une porteuse, à phase initiale aléatoire, modulée en amplitude et en phase. Le dispositif de détection (de discrimination) se compose de filtres généralisés, adaptés à chaque quadrature du signal, de détecteurs quadratiques, de sommateurs et d'un dispositif à seuil (dispositif

*) Le premier paramètre entre parenthèses est égal à la valeur moyenne de la grandeur aléatoire, et le second à sa variance.

de comparaison). Une variante plus simple se compose d'un seul filtre adapté et d'un détecteur linéaire par signal.

La structure du détecteur d'un signal quasi déterministe du type général (1.9) en présence d'un bruit additif normal corrélé de matrice de corrélation M est suggérée par la moyenne du rapport de vraisemblance, qui dans le cas envisagé est de la forme

$$\Lambda(\mathbf{x}) = \int_{\Theta} l(\mathbf{x}|\boldsymbol{\vartheta})\, w_m(\boldsymbol{\vartheta})\, d\boldsymbol{\vartheta}, \tag{1.31}$$

où $l(\mathbf{x} \mid \boldsymbol{\vartheta}) \equiv l(\mathbf{x} \mid \mathbf{s}) = \exp(\mathbf{x}'\mathbf{M}^{-1}\mathbf{s} - 0{,}5\mathbf{s}'\mathbf{M}^{-1}\mathbf{s})$.

Après la substitution $s_j = s(t_j) = \boldsymbol{\vartheta}'\boldsymbol{\varphi}(t_j)$ on obtient en vertu de (1.9')

$$l(\mathbf{x}|\boldsymbol{\vartheta}) = \exp\left(\boldsymbol{\vartheta}'\mathbf{y} - \frac{1}{2}\boldsymbol{\vartheta}'\mathbf{B}\boldsymbol{\vartheta}\right), \tag{1.32}$$

où

$$\mathbf{y} = \mathbf{M}^{-1}\mathbf{x}\boldsymbol{\Phi}', \tag{1.32'}$$

$$\mathbf{B} = \boldsymbol{\Phi}'\mathbf{M}^{-1}\boldsymbol{\Phi}, \tag{1.32''}$$

$\boldsymbol{\Phi}$ étant une matrice $m \times n$ à éléments $\varphi_i(t_j)$ et

$$\mathbf{x}'\mathbf{M}^{-1}\mathbf{s} = \mathbf{x}'\mathbf{M}^{-1}\boldsymbol{\Phi}'\boldsymbol{\vartheta} = \mathbf{y}'\boldsymbol{\vartheta} = \boldsymbol{\vartheta}'\mathbf{y},$$

$$\mathbf{s}'\mathbf{M}^{-1}\mathbf{s} = \boldsymbol{\vartheta}'\boldsymbol{\Phi}\mathbf{M}^{-1}\boldsymbol{\Phi}'\boldsymbol{\vartheta} = \boldsymbol{\vartheta}'\mathbf{B}\boldsymbol{\vartheta}.$$

La matrice de corrélation M est auto-adjointe (hermitienne), par conséquent $(\mathbf{M}^{-1})' = \mathbf{M}^{-1}$, $\mathbf{B} = \mathbf{B}'$. Si la distribution a priori des paramètres du signal (1.9) est normale de vecteur des moyennes $\boldsymbol{\vartheta}_0$ et de matrice de corrélation D, en calculant l'intégrale (1.31) comme convolution des distributions normales on obtient

$$\Lambda(\mathbf{x}) = \left[\frac{\det \mathbf{D}}{\det(\mathbf{B}+\mathbf{D}^{-1})}\right]^{1/2} \times$$

$$\times \exp\left[\frac{1}{2}\mathbf{y}'\mathbf{B}^{-1}\mathbf{y} - \frac{1}{2}(\mathbf{y} - \mathbf{B}\boldsymbol{\vartheta}_0)'(\mathbf{B} + \mathbf{B}'\mathbf{D}^{-1}\mathbf{B})^{-1}(\mathbf{y} - \mathbf{B}\boldsymbol{\vartheta}_0)\right]. \tag{1.33}$$

Le détecteur optimal d'un *signal stochastique*, représentant un processus aléatoire normal, dans un mélange additif avec un bruit blanc se compose d'un filtre à paramètres variables, d'une ligne à retard, d'un multiplicateur, d'un intégrateur et d'un dispositif à seuil. L'algorithme de détection est donné par la formule

$$\int_{-T}^{T}\int_{-T}^{T} x(t)\, x(v)\, h(t, v)\, dt\, dv \geqslant C \tag{1.34}$$

(voir tome II, 5.2.3). La caractéristique $h(t, v)$ du filtre s'obtient à partir d'une équation intégrale linéaire et dépend de la fonction de corrélation $B_s(t, y)$ du signal et de la densité spectrale $N_0$ du

bruit blanc

$$\int_{-T}^{T} B_s(t, y)\, h(y, v)\, dy + N_0 h(t, v) =$$
$$= B_s(t, v), \quad |t| \leqslant T, \quad |v| \leqslant T. \tag{1.34'}$$

Bien que les algorithmes mentionnés correspondent au cas particulier de détection des signaux noyés dans un bruit normal additif, leur valeur dépasse le cadre restreint du cas particulier. Comme nous allons le voir plus loin, certaines méthodes utilisées pour surmonter l'indétermination a priori ne rejettent pas les résultats obtenus dans l'hypothèse d'une distribution normale du bruit, mais s'appuient sur ces résultats.

**1.1.10. Estimation des paramètres et filtrage.** L'estimation optimale (de Bayes) $\hat{\boldsymbol{\vartheta}}$ du paramètre vectoriel $\boldsymbol{\vartheta}$ d'un signal est égale à la moyenne conditionnelle de ce paramètre, si toutefois la densité a posteriori du paramètre est symétrique par rapport au mode et si la fonction de pertes est paire

$$\hat{\boldsymbol{\vartheta}} = \int_{\Theta} \boldsymbol{\vartheta} W(\boldsymbol{\vartheta} | \mathbf{x})\, d\boldsymbol{\vartheta}, \tag{1.35}$$

où

$$W(\boldsymbol{\vartheta} | \mathbf{x}) = \frac{w_m(\boldsymbol{\vartheta})\, W(\mathbf{x} \mid \boldsymbol{\vartheta})}{\int_{\Theta} w_m(\boldsymbol{\vartheta})\, W(\mathbf{x} \mid \boldsymbol{\vartheta})\, d\boldsymbol{\vartheta}}. \tag{1.36}$$

Pour un signal quasi déterministe de la forme (1.9) les estimations optimales sont données par le système d'équations suivant (algorithme discret) [2], dans l'hypothèse que son paramètre vectoriel est normal, en présence d'un bruit blanc additif normal corrélé et pour une fonction de pertes quadratique :

$$\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0 = \mathbf{C}[\mathbf{z} + \mathbf{A}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0)], \tag{1.37}$$

où $\mathbf{C}^{-1} = \mathbf{A} + \mathbf{B}^{-1}$,

$$z_k = \frac{\partial \ln L}{\partial \vartheta_k}\bigg|_{\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0}, \tag{1.37'}$$

$$A_{kj} = \frac{\partial^2 \ln L}{\partial \vartheta_k\, \partial \vartheta_j}\bigg|_{\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0}, \tag{1.37''}$$

où $\boldsymbol{\vartheta}_0$ et $\mathbf{B}$ sont respectivement le vecteur moyenne et la matrice des covariances de la distribution normale du paramètre estimé $\boldsymbol{\vartheta}$; $L \equiv L_{\mathbf{x}}(\boldsymbol{\vartheta})$ est la fonction de vraisemblance du vecteur des observations $\mathbf{x} = (x_1, \ldots, x_n)$.

Lorsque l'on utilise des réalisations continues, le système d'équations (1.37) devient une équation intégrale linéaire que l'on peut

interpréter à l'aide d'un mesureur asservi formé par des dispositifs de traitement non linéaire du signal d'entrée, un filtre de lissage et un additionneur. La théorie quasi linéaire de Bolchakov-Répine [1] conduit également à l'algorithme (1.37). Cette théorie part de l'approximation normale de la distribution multidimensionnelle a posteriori du paramètre vectoriel estimé en s'appuyant sur la condition de haute précision a posteriori des estimations.

Lors du *filtrage linéaire optimal* on se limite à des algorithmes linéaires et on utilise généralement le critère du minimum du carré moyen de l'erreur. La caractéristique d'un filtre linéaire s'obtient à partir de l'équation de Wiener-Hopf. Pour trouver un algorithme discret de filtrage optimal on utilise le modèle discret de cette équation. Lorsque l'on utilise des algorithmes linéaires, l'optimum absolu ne peut être obtenu que si le signal et le bruit sont répartis suivant une loi normale et qu'ils sont additifs. R. Kalman et R. Bucy [11] ont obtenu des résultats importants sous forme récurrente pour le cas non stationnaire.

**1.1.11. Détection simultanée et estimation des paramètres des signaux.** Pour une fonction de pertes quadratique, la relation entre l'estimation de Bayes $\hat{\vartheta}_{p=1}$ du paramètre $\vartheta$ obtenue dans l'hypothèse que le signal est présent dans le processus observé, et l'estimation de Bayes $\hat{\vartheta}_{p<1}$ de ce paramètre dans l'incertitude quant à la présence du signal, est donnée par une relation obtenue par D. Middleton et R. Esposito [7]

$$\hat{\vartheta}_{p<1} = \frac{\Lambda}{\Lambda+1}\,\hat{\vartheta}_{p=1}, \tag{1.38}$$

$$\Lambda = \frac{p}{1-p}\,\frac{1}{W(\mathbf{x}\mid 0)}\int_{-\infty}^{\infty} w_1(\vartheta)\,W(\mathbf{x}\mid\vartheta)\,d\vartheta. \tag{1.38'}$$

Ces résultats peuvent être généralisés au cas des alternatives multiples et des fonctions de pertes arbitraires.

Pour des processus aléatoires markoviens le problème de la détection simultanée et du filtrage a été posé et résolu par R. Stratonovitch et Yu. Sossouline [8]. Ils ont établi la relation suivante entre la fonctionnelle du rapport de vraisemblance et l'estimation de Markov $\hat{s}(t)$ du signal d'après le critère du minimum du carré moyen de l'erreur dans le problème de la détection d'un signal markovien en présence d'un bruit blanc normal additif:

$$\ln l[x(t)] = \frac{1}{N_0}\left[\int_0^T \hat{s}(t)\,x(t)\,dt - \frac{1}{2}\int_0^T [\hat{s}(t)]^2\,dt\right], \tag{1.39}$$

où $N_0$ est la densité spectrale du bruit blanc; $x(t)$ la réalisation observée sur l'intervalle $(0, T)$ donné. T. Kailath [3] a généralisé d'une manière très élégante ces résultats, utilisant à cet effet la méthode du processus d'innovation.

**1.1.12. Résumé.** Dans son ensemble, l'aperçu que nous venons de donner reflète l'état de la théorie de la synthèse statistique des systèmes informationnels vers la fin des années 60. Pour obtenir des solutions générales on a besoin d'un grand nombre de données a priori. Les solutions particulières sous forme fermée, assez facilement interprétables, ont été obtenues pour une classe restreinte de modèles des signaux et des bruits. La présence de bruits corrélés entraîne toujours de grandes difficultés de calcul liées à l'inversion de matrices d'ordre élevé ou à la résolution d'équations intégrales. Tout changement de modèle entraîne un changement fondamental de l'algorithme. La critique de cet état de choses dans la théorie a donné naissance à une nouvelle branche, liée à l'élaboration de méthodes permettant de surmonter l'indétermination a priori dans les problèmes de la synthèse statistique, à la recherche d'algorithmes voisins de l'optimum (optimum pour une information a priori complète) et stables, ne serait-ce que partiellement, vis-à-vis des variations éventuelles de la conjoncture initiale.

## 1.2 PROBLÈMES DE LA SYNTHÈSE DANS LES CONDITIONS DE L'INDÉTERMINATION A PRIORI

**1.2.1. Indétermination a priori paramétrique et non paramétrique.** On distingue deux classes de problèmes de la synthèse statistique des systèmes d'information dans des conditions d'indétermination a priori: paramétriques et non paramétriques.

Dans les problèmes à indétermination a priori *paramétrique* on donne une famille paramétrique de fonctions de vraisemblance $w_n(\mathbf{x} \mid \boldsymbol{\vartheta})$ d'un échantillon ou un ensemble paramétrique de fonctionnelles du rapport de vraisemblance $l[x(t) \mid \boldsymbol{\vartheta}]$ d'une réalisation. L'indétermination a priori concerne certains paramètres ou tout l'ensemble de paramètres $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_n)$ dont dépendent les caractéristiques du signal et du bruit.

Si le vecteur des paramètres $\boldsymbol{\vartheta}$ est donné, les observations qui sont un échantillon aléatoire ou une réalisation d'un processus aléatoire, sont entièrement déterminées du point de vue statistique. Cependant, les cas où les familles $w_n(\mathbf{x} \mid \boldsymbol{\vartheta})$ (ou $l[x(t) \mid \boldsymbol{\vartheta}]$ ) sont constituées par un seul élément et par conséquent les hypothèses vérifiées sont simples et sont plutôt d'heureuses exceptions; plus souvent on a affaire à une famille de distributions paramétrique pour laquelle le paramètre vectoriel $\boldsymbol{\vartheta}$ appartient à un certain domaine de l'espace euclidien à $m$ dimensions (ou à l'espace tout entier).

Il est naturel de demander quelle est la nature de ce paramètre: est-ce une *constante* vectorielle inconnue ou une grandeur vectorielle *aléatoire*. Autrement dit, y a-t-il sur l'ensemble $\Theta$ une densité $w_m(\vartheta)$ qui détermine les valeurs plus ou moins probables de $\vartheta$ ou bien cette caractéristique probabiliste n'existe pas? La réponse à cette question constitue également une donnée a priori, et chacune des deux réponses mentionnées peut être mise à la base d'une conception qui permettra de formuler les critères de qualité et les méthodes de résolution des problèmes de la synthèse statistique.

La théorie de Bayes de vérification des hypothèses statistiques et d'estimation des paramètres est fondée sur la *conception de paramètres aléatoires* à densité mutuelle $w_m(\vartheta)$ donnée *). Dans les conditions d'indétermination a priori, les problèmes correspondants se formulent dans l'hypothèse que le paramètre vectoriel $\vartheta$ est aléatoire et que l'indétermination a priori concerne la densité $w_m(\vartheta)$ de ce paramètre.

Dans les problèmes d'indétermination a priori *non paramétrique*, la famille des fonctions de vraisemblance $w_n(\mathbf{x})$ de l'échantillon est non paramétrique en ce sens qu'on ne peut pas lui faire correspondre d'une manière univoque un certain domaine de l'espace euclidien de dimension finie. Une famille non paramétrique de fonctions de vraisemblance peut comprendre, par exemple, toutes les fonctions normées non négatives.

Par exemple, la famille de fonctions de vraisemblance des bruits normaux corrélés, à fonction de corrélation inconnue (ou à spectre énergétique inconnu) est non paramétrique. Dans le sens de Bayes, la famille de distributions $w_m(\vartheta)$ des paramètres inconnus $\vartheta = (\vartheta_1, \ldots, \vartheta_m)$ de la fonction de vraisemblance d'un échantillon peut être une famille non paramétrique.

La théorie non paramétrique des décisions est développée essentiellement pour les problèmes à deux hypothèses; à titre d'exemple on peut citer le problème de la détection d'un signal en présence de bruits. On entend par *algorithmes non paramétriques* les algorithmes de traitement des observations invariants (c'est-à-dire laissant constante la valeur de la probabilité d'erreur) par rapport à l'une des *hypothèses non paramétriques*. Par exemple, dans le problème de la détection d'un signal noyé dans un bruit l'utilisation d'un algorithme non paramétrique assure la constance de la probabilité de fausse alarme pour une répartition quelconque des probabilités des bruits. Si l'invariance de l'algorithme est assurée à la limite, lorsque la taille de l'échantillon tend vers l'infini, l'algorithme est dit *asymptotiquement non paramétrique*. Un algorithme « réellement » non para-

*) Dans la théorie de Bayes le moyennage des fonctions de vraisemblance sur les paramètres aléatoires permet de réduire le problème de la vérification des hypothèses composées à la vérification des hypothèses simples.

métrique devrait être invariant par rapport aux deux hypothèses en vérification. L'établissement des *estimations* de fonctions de vraisemblance $w_n$ (x) inconnues d'après un échantillon observé est également un problème de la statistique non paramétrique.

L'*invariance partielle* de l'algorithme de traitement des observations vis-à-vis de l'une des hypothèses non paramétriques est parfois intéressante. Ainsi, lorsqu'il s'agit de détecter un signal en présence de bruits, on peut formuler le problème de la synthèse statistique d'un détecteur, optimal du point de vue d'un certain critère de qualité pour une distribution arbitraire des bruits et tel qu'une partie de sa structure est réalisée pour une distribution donnée des bruits, et l'autre reste invariante et ne dépend pas de la distribution des bruits.

**1.2.2. Algorithmes adaptatifs.** Pour surmonter l'indétermination a priori on utilise largement la méthode d'*apprentissage* préalable, c'est-à-dire l'élaboration, à partir des observations, d'estimations des fonctions de répartition inconnues, à condition de disposer d'une famille paramétrique de fonctions de répartition. Ces estimations sont ensuite utilisées à la place des caractéristiques probabilistes inconnues des processus étudiés.

On distingue l'*apprentissage avec maître*, permettant d'obtenir une suite d'observations étalons, on sait de plus à quelle distribution appartient chacun des éléments (échantillon d'apprentissage *classifié*), et l'*apprentissage sans maître* (auto-apprentissage) où les estimations mentionnées sont formées à partir d'échantillons d'apprentissage *non classifiés*.

Dans certains cas les échantillons d'apprentissage peuvent être utilisés pour l'estimation des caractéristiques de l'algorithme (par exemple, dans le problème de la détection d'un signal pour l'estimation d'un seuil ou pour l'estimation de la forme de transformation fonctionnelle de certaines statistiques).

Nous appellerons algorithmes *adaptatifs* les algorithmes de traitement des observations *) utilisant les estimations des fonctions de répartition, de leurs paramètres ou d'autres caractéristiques quelconques obtenues par suite de l'apprentissage.

**1.2.3. Critères de qualité des algorithmes de détection des signaux dans les conditions d'indétermination paramétrique.** Lors de la vérification d'une hypothèse simple contre l'alternative simple, la comparaison du rapport de vraisemblance avec un seuil est un algorithme assurant une probabilité donnée de fausse alarme et une pro-

*) Parfois le terme adaptatif est utilisé dans le sens de l'invariance par rapport aux changements de situation. Par adaptation on entend ici la variation des paramètres et de la structure du système (algorithme de prise de décision) par suite de l'apprentissage.

babilité minimale de perte de signal (critère de Neyman-Pearson). La forme la plus simple d'une indétermination paramétrique, un signal constant pouvant prendre une valeur réelle quelconque, correspond à une alternative *composite*. Comme nous l'avons mentionné dans le second tome, dans la plupart des cas (1.4.7), la règle de détection uniformément la plus puissante, minimisant la probabilité de perte de signal, n'existe pas. Par exemple, lors de la détection d'après un échantillon indépendant d'un signal constant en présence d'un bruit normal stationnaire additif de moyenne nulle et de variance donnée, la comparaison de la somme des valeurs échantillonnées avec un seuil donne la règle de détection uniformément la plus puissante seulement si l'on suppose que le signe du signal ne change pas. Sinon on devra se limiter à la classe de règles *non biaisées*, en introduisant la condition naturelle que la probabilité de détection (quelle que soit la valeur du signal) soit non inférieure à la probabilité de fausse alarme.

Dans le cas où l'hypothèse de l'absence de signal est elle aussi composite, c'est-à-dire qu'il y a indétermination paramétrique quant au bruit, la probabilité de fausse alarme dépend également des paramètres de la distribution des bruits. Par exemple, si la variance d'un bruit normal stationnaire additif (de moyenne nulle) est inconnue, la probabilité de fausse alarme dépend de la variance du bruit, qui est une constante positive arbitraire. Dans ce cas on peut essayer de trouver une règle *similaire*, telle que la probabilité de fausse alarme ne dépende pas de la variance $\sigma^2$ du bruit, c'est-à-dire

$$\int_{X_1} w_n(\mathbf{x}|\sigma^2)\,d\mathbf{x} = \alpha, \tag{1.40}$$

où $w_n(\mathbf{x} \mid \sigma^2)$ est la fonction de vraisemblance de l'échantillon du bruit et $X_1$ le domaine critique correspondant à la décision sur la présence du signal.

Si le domaine critique $X_1$ satisfaisant à la condition (1.40) existe, on dit qu'il est semblable à l'espace des échantillons, car pour tout l'espace des échantillons on a toujours

$$\int_{X} w_n(\mathbf{x}|\sigma^2)\,d\mathbf{x} = 1. \tag{1.40'}$$

Cette intégrale ne dépend non plus du paramètre inconnu.

Dans le choix d'une règle statistique de vérification d'une hypothèse composite contre une alternative composite, on peut parfois partir des considérations de son *invariance* par rapport à un certain *groupe de transformations* des coordonnées de l'espace échantillonné. Ainsi, par exemple, si la règle utilise les statistiques

$y = (x_1 - x_n, \ldots, x_{n-1} - x_n)$ ou $y = (x_1 - m_1^*, \ldots, x_n - m_1^*)$, où

$$m_1^* = \frac{1}{n} \sum_{k=1}^{n} x_k,$$

elle sera invariante par rapport au *groupe des déplacements*, mais si l'on utilise la statistique

$$y = \sum_{k=1}^{n} x_i^2,$$

elle sera invariante par rapport au *groupe des rotations*. Chaque fois où l'on se limite à la classe de règles invariantes par rapport à un groupe donné de transformations, le problème de la synthèse optimale revient à la détermination de la règle invariante uniformément la plus puissante ou de la règle invariante non biaisée uniformément la plus puissante.

Enfin, dans la conception non bayessienne du choix du critère de qualité de l'algorithme de détection des signaux, dans les conditions d'indétermination paramétrique, il y a lieu de mentionner également le critère du minimax (1.1.4), bien qu'il soit parfois difficile de rechercher les règles du minimax.

Soulignons que les critères énumérés, à savoir l'absence de biais, la similitude, l'invariance, le minimax, se complètent mutuellement.

Dans certains cas, pour tester une hypothèse composite $H$ selon laquelle le paramètre $\vartheta$ appartient au domaine $\Theta_H$, contre l'alternative $K$, selon laquelle il appartient au domaine $\Theta_K$ on utilise le *critère du maximum de vraisemblance*: l'hypothèse $H$ est rejetée si

$$\sup_{\vartheta \in \Theta_K} w_n(\mathbf{x}|\vartheta) \geqslant \sup_{\vartheta \in \Theta_H} w_n(\mathbf{x}|\vartheta), \tag{1.41}$$

où sous une forme plus générale

$$\sup_{\vartheta \in \Theta_K} w_n(\mathbf{x}|\vartheta) / \sup_{\vartheta \in \Theta_H} w_n(\mathbf{x}|\vartheta) \geqslant C. \tag{1.41'}$$

Si le paramètre $\vartheta$ est aléatoire et l'on connaisse ses distributions $w_H(\vartheta)$ pour l'hypothèse et $w_K(\vartheta)$ pour l'alternative, le moyennage de $w_n(\mathbf{x} \mid \vartheta)$ pour chacune de ces distributions permet de réduire le problème aux hypothèses composites à un problème de la vérification des hypothèses simples (voir tome II, 1.4.7).

**1.2.4. Critères de qualité des algorithmes de détection des signaux dans les conditions d'indétermination non paramétrique.** Nous allons maintenant envisager le cas où la famille de fonctions de vraisemblance de l'échantillon est non paramétrique. L'hypothèse $H$: il n'y a pas de signal, et l'alternative $K$: le signal est présent, sont composites. Désignons par $\mathbf{W}$ la famille non paramétrique de fonc-

tions de vraisemblance de l'échantillon $\mathbf{x} = (x_1, \ldots, x_n)$ et par $\mathbf{X}_k$ le domaine critique de l'espace des échantillons. Lorsque l'échantillon tombe dans le domaine $\mathbf{X}_k$ on prend la décision que le signal est présent. La caractéristique de travail du détecteur (fonction de puissance de la règle du choix de la décision) dans les conditions envisagées de l'indétermination non paramétrique est la probabilité pour l'échantillon $\mathbf{x}$ d'appartenir au domaine $\mathbf{X}_k$ lorsque la fonction de vraisemblance de l'échantillon est égale à une certaine fonction $w_n(\mathbf{x})$ appartenant à la famille W. Cette probabilité est égale à la probabilité de détection correcte, lorsque $w_n(\mathbf{x})$ est la distribution du mélange du signal et du bruit, et égale à la probabilité de fausse alarme lorsque $w_n(\mathbf{x})$ est la distribution du bruit, c'est-à-dire

$$P\{\mathbf{x} \in \mathbf{X}_k \mid w_n(\mathbf{x}), K\} = 1 - \beta\{w_n\}, \tag{1.42}$$

si $w_n$ correspond à l'alternative $K$ et

$$P\{\mathbf{x} \in \mathbf{X}_k \mid w_n(\mathbf{x}), H\} = \alpha\{w_n\}, \tag{1.42'}$$

si $w_n$ correspond à l'hypothèse $H$.

Nous allons nous limiter à la classe $\Delta_\alpha$ des algorithmes de détection consistants non paramétriques pour lesquels la probabilité de fausse alarme $\alpha$ est constante quelles que soient les distributions des bruits

$$P\{\mathbf{x} \in \mathbf{X}_k \mid w_n(\mathbf{x})\} = \alpha \tag{1.43}$$

pour toutes $w_n(\mathbf{x})$ correspondant à l'hypothèse $H$. Nous allons nous donner de plus une certaine valeur $\beta$ de la probabilité de perte du signal. Pour une alternative $K$ donnée, considérons deux suites d'algorithmes de détection $\delta_n$ et $\delta^*_{n*}$ appartenant à la classe $\Delta_\alpha$, où $\{n\}$ et $\{n^*\}$ sont les suites de tailles des échantillons. En vertu de la consistance, pour chacune de ces suites d'algorithmes il y aura toujours des tailles minimales $n_k$, $n^*_k$ des échantillons telles que

$$\beta\{w_{n_k}, \delta_{n_k}\} \leqslant \beta, \quad \beta\{w_{n^*_k}, \delta^*_{n^*_k}\} \leqslant \beta. \tag{1.44}$$

La qualité de l'algorithme de détection du signal dans les conditions d'indétermination non paramétrique est définie par la notion d'*efficacité relative*. L'algorithme non paramétrique $\delta^*_n$ de la classe $\Delta_\alpha$ est plus efficace que l'algorithme non paramétrique $\delta_n$ de la même classe, si avec la condition (1.44) on a $n^*_k \leqslant n_k$. Autrement dit, plus efficace des deux algorithmes non paramétriques maintenant constant le niveau des fausses alarmes sera celui pour lequel la probabilité donnée de détection correcte est atteinte pour un échantillon de taille plus réduite. On utilise comme mesure d'efficacité le rapport des tailles mentionnées de l'échantillon que l'on appelle *coefficient d'efficacité relative* de l'algorithme $\delta^*_{n*}$ par rapport à l'algorithme $\delta_n$

$$\rho_{\alpha,\beta}(\delta_n, \delta^*_{n*}) = n_k/n^*_k. \tag{1.45}$$

L'algorithme $\delta_{n*}^{*}$ est plus efficace que $\delta_n$ si

$$\rho_{\alpha,\beta}(\delta_n, \delta_{n*}^{*}) \geqslant 1.$$

Supposons que pour des distributions données du bruit et du signal avec bruit, l'algorithme $\delta_n^0$ est optimal d'après le critère de Neyman-Pearson, c'est-à-dire qu'il assure le minimum de probabilité de perte du signal pour une probabilité donnée de fausse alarme. Il est évident que le coefficient d'efficacité relative de l'algorithme non paramétrique $\delta_{n*}^{*}$ par rapport à $\delta_n^0$ pour les distributions données sera inférieur à l'unité (ou au moins égal à l'unité). Cependant, pour une autre distribution du bruit, l'algorithme non paramétrique $\delta_{n*}^{*}$ peut s'avérer être plus efficace que l'algorithme $\delta_n^0$, optimal en présence d'un bruit du type donné.

**1.2.5. Critères de qualité des algorithmes adaptatifs.** Le critère général de qualité des algorithmes adaptatifs est leur convergence vers les algorithmes optimaux correspondants à information a priori complète lorsque la taille des échantillons d'apprentissage augmente indéfiniment. Les algorithmes adaptatifs, en tant que statistiques des échantillons d'apprentissage, convergeant en probabilité vers les algorithmes optimaux lorsque la taille de l'échantillon d'apprentissage augmente indéfiniment, sont appelés algorithmes adaptatifs *consistants*. La condition suffisante de consistance d'un algorithme adaptatif est sa convergence en moyenne quadratique vers l'algorithme optimal.

Il est évident que tout algorithme adaptatif ne doit pas être consistant, car l'apprentissage sur des échantillons aléatoires (classifiés ou non) introduit un élément aléatoire supplémentaire.

Dans les problèmes à indétermination paramétrique, en particulier dans les problèmes à paramètres perturbateurs, on construit des algorithmes adaptatifs à partir des algorithmes optimaux (d'après un critère de qualité de Bayes ou autre, en remplaçant les paramètres inconnus, comprenant en particulier des seuils inconnus, par leurs estimations obtenues à l'aide d'échantillons d'apprentissage *). Lorsqu'il s'agit d'une détection ou d'une séparation des signaux en présence de paramètres perturbateurs (bruit), on construit des estimations à l'aide d'un échantillon d'apprentissage du bruit. Dans le problème de *classification*, où il y a lieu de déterminer l'algorithme de traitement d'observations qui permette de rapporter chaque

*) Remarquons que l'algorithme (1.41′) peut être interprété comme une variante adaptative prescrivant une comparaison du rapport de vraisemblance avec un seuil. En effet, la condition sup $w_n(\mathbf{x} \mid \boldsymbol{\vartheta})$ est satisfaite par une certaine statistique $\hat{\boldsymbol{\vartheta}}(\mathbf{x})$ qui peut être adoptée en qualité d'estimation du paramètre inconnu $\boldsymbol{\vartheta}$. L'inégalité (1.41′) peut alors s'écrire sous la forme

$$w_n(\mathbf{x} \mid \hat{\boldsymbol{\vartheta}}_K)/w_n(\mathbf{x} \mid \hat{\boldsymbol{\vartheta}}_H) \geqslant C.$$

observation à l'un des ensembles décrit du point de vue probabiliste d'une manière incomplète, on utilise l'apprentissage d'après des échantillons d'apprentissage classifiés pour chacun des ensembles ou l'auto-apprentissage d'après un échantillon d'observations non classifié.

Il y a lieu de noter que le critère de consistance des algorithmes adaptatifs ne permet pas de définir d'une manière univoque le type des estimations des paramètres inconnus. Souvent on prend pour ces estimations les estimations du maximum de vraisemblance (voir § 4.5).

Considérons à titre d'exemple le problème de la détection du signal $\lambda s(t)$ noyé dans un bruit, dont la distribution $w(\mathbf{x}, \boldsymbol{\vartheta})$ dépend du paramètre vectoriel (perturbateur) $\boldsymbol{\vartheta}$. L'algorithme de détection, optimal d'après le critère de Neyman-Pearson, préscrit la comparaison du rapport de vraisemblance $l(\mathbf{x} \mid \lambda s, \boldsymbol{\vartheta})$ avec le seuil $C_\alpha(\boldsymbol{\vartheta})$. L'algorithme adaptatif correspondant s'obtient en remplaçant l'inconnue $\boldsymbol{\vartheta}$ par l'estimation $\hat{\boldsymbol{\vartheta}}$. Cet algorithme adaptatif sera consistant si, lorsque l'estimation $\hat{\boldsymbol{\vartheta}}$ est consistante, la statistique $l(\mathbf{x} \mid \lambda s, \hat{\boldsymbol{\vartheta}})/C_\alpha(\hat{\boldsymbol{\vartheta}})$ converge en probabilité vers $l(\mathbf{x} \mid \lambda s, \boldsymbol{\vartheta})/C_\alpha(\boldsymbol{\vartheta})$ lorsque la taille de l'échantillon utilisé pour l'estimation croît indéfiniment.

Si y est un échantillon d'apprentissage classifié appartenant à la distribution du bruit et $\hat{\boldsymbol{\vartheta}}(\mathbf{y})$ une estimation consistante de $\boldsymbol{\vartheta}$, la condition suffisante de consistance de l'algorithme adaptatif

$$l(\mathbf{x} \mid \lambda s, \hat{\boldsymbol{\vartheta}}) \geqslant C_\alpha(\hat{\boldsymbol{\vartheta}}) \tag{1.46}$$

est la continuité de la statistique $l(\mathbf{x} \mid \lambda s, \boldsymbol{\vartheta})/C_\alpha(\boldsymbol{\vartheta})$ en $\boldsymbol{\vartheta}$ pour tous les $\mathbf{x}$. Une autre condition de consistance est que la variance de la statistique $l(\mathbf{x} \mid \lambda s, \boldsymbol{\vartheta})$ soit limitée.

En auto-apprentissage, lorsque l'estimation s'effectue d'après l'échantillon observé $\mathbf{x} = (x_1, \ldots, x_n)$, l'algorithme adaptatif peut se trouver être non consistant même pour une estimation consistante lorsque l'hypothèse $H$ est vérifiée, car pour l'alternative cette estimation peut converger en probabilité vers $\boldsymbol{\vartheta}_1 \neq \boldsymbol{\vartheta}$.

Dans les problèmes à indétermination non paramétrique les algorithmes adaptatifs sont construits à l'aide d'estimations obtenues d'après des échantillons d'apprentissage de fonctions de répartition inconnues, de fonctions de corrélation, de spectres énergétiques ou d'autres fonctions inconnues.

**1.2.6. Critères de qualité asymptotiques.** L'utilisation lors de la synthèse statistique dans des conditions d'indétermination a priori des critères de qualité envisagés se heurte souvent à des difficultés insurmontables. Il n'existe pas à l'heure actuelle de méthodes régulières suffisamment simples de construction d'algorithmes uniformé-

ment les plus puissants non biaisés, semblables, invariants ou à minimax. Le calcul sous forme fermée du coefficient d'efficacité relative, dépendant des grandeurs données $\alpha$, $\beta$ et de la distribution du signal noyé dans le bruit, est dans la plupart des cas impossible à réaliser. Les lois de distribution des statistiques définissant les algorithmes adaptatifs pour des tailles finies des échantillons d'apprentissage et d'observation sont très compliquées; il est par conséquent impossible de caractériser dans ces conditions l'algorithme adaptatif par une mesure quantitative.

L'histoire des sciences connaît pas mal d'exemples où les difficultés accompagnant certains problèmes ont été surmontées par la recherche de solutions asymptotiques. Les théorèmes de Shannon — théorèmes fondamentaux de la théorie de l'information — sont, comme on sait, asymptotiques. En théorie des probabilités et en mathématique statistique les méthodes asymptotiques ont été utilisées depuis Laplace et Gauss et se trouvent à la base de la loi des grands nombres et de nombreux théorèmes limites. Il est naturel de faire appel, dans un certain nombre de cas, aux méthodes asymptotiques d'étude des problèmes de la synthèse statistique dans les conditions d'indétermination a priori. L'utilisation de ces méthodes est basée sur le choix d'un certain système de paramètres du problème dont la solution, obtenue dans l'hypothèse que ces paramètres augmentent indéfiniment ou tendent vers une certaine limite finie, définit un algorithme aussi voisin que l'on veut de l'optimal dans les conditions d'information a priori complète. Comme, en réalité, les algorithmes sont utilisés dans les cas où les paramètres n'atteignent pas leur limite, il y a toujours lieu d'étudier la vitesse de convergence de l'algorithme asymptotiquement optimal vers l'algorithme optimal.

Rappelons que pour les algorithmes adaptatifs on utilise toujours des critères de qualité asymptotiques pour un temps d'apprentissage infini (tailles des échantillons d'apprentissage infiniment grandes).

Dans de nombreux cas, lorsque les statistiques utilisées pour prendre des décisions ou trouver des estimations peuvent être présentées sous la forme de sommes de grandeurs aléatoires, on arrive à formuler les conditions où le théorème de la limite centrale est applicable et peut être utilisé pour la recherche des caractéristiques asymptotiques des statistiques mentionnées.

**1.2.7. Algorithmes localement optimaux de détection des signaux.** Considérons une famille paramétrique de fonctions de vraisemblance $w_n(\mathbf{x} \mid \vartheta)$ d'un échantillon de taille fixe $n$ du mélange d'un signal et d'un bruit avec paramètre énergétique $\vartheta$ (par exemple, le rapport de la puissance du signal à la puissance du bruit). Le terme « énergétique » signifie que pour $\vartheta = 0$ le signal disparaît et la distribution du mélange du signal et du bruit $w_n(\mathbf{x} \mid \vartheta)$ devient la distribution

du bruit $w_n$ (x | 0) (i.e. on suppose que l'interaction du signal et du bruit est telle que pour $\vartheta = 0$ le bruit ne disparaît pas).

Nous dirons qu'un algorithme de détection est *localement optimal* ($\varepsilon$-optimal) si, pour une probabilité donnée de fausse alarme $\alpha$, il assure une probabilité minimale $\beta(\vartheta)$ de perte du signal dans un certain intervalle $0 < \vartheta < \varepsilon$ des valeurs de $\vartheta$. Il est évident que l'algorithme uniformément le plus puissant est localement optimal, la réciproque n'étant pas vraie.

L'idée d'un algorithme localement optimal est illustrée par la figure 1.1 où l'on a représenté, en fonction du rapport signal/bruit, la caractéristique de travail de détection (fonction de puissance

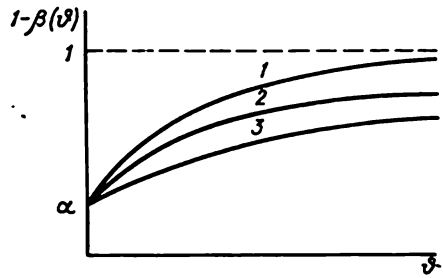



Fig. 1.1. Caractéristiques de travail des algorithmes de détection: non biaisé uniformément le plus puissant (1), localement optimal (2) et non optimal (3)-

$1 - \beta(\vartheta)$) dans les cas d'un algorithme uniformément le plus puissant non biaisé (courbe 1), d'un algorithme localement optimal (courbe 2) et d'un algorithme non optimal (courbe 3) pour une probabilité de fausse alarme donnée $\alpha = 1 - \beta(0)$. Si $\varepsilon$ est petit, l'algorithme localement optimal se rapporte au cas d'un signal faible détecté pour un temps d'observation fini (pour une taille finie de l'échantillon). Dans ce cas, on exige que les quantités $\frac{\partial}{\partial\vartheta}\beta(\vartheta)|_{\vartheta=0}$ coïncident pour les algorithmes optimal $\delta_n^0$ et localement optimal $\delta_n^{lo}$ et que pour tout autre algorithme on ait

$$\frac{\partial}{\partial\vartheta}\beta(\vartheta, \delta_n)|_{\vartheta=0} > \frac{\partial}{\partial\vartheta}\beta(\vartheta, \delta_n^{lo})|_{\vartheta=0}.$$

Pour trouver l'algorithme de détection localement optimal on développe en série de Taylor le logarithme du rapport de vraisemblance (ou de la moyenne du rapport de vraisemblance, ou d'une fonctionnelle du rapport de vraisemblance) au voisinage du point $\vartheta = 0$ et on se limite au terme linéaire en $\vartheta$. Le choix d'une décision est soumis à la règle suivante: le signal est présent si

$$\frac{\partial}{\partial\vartheta}\ln l(\mathbf{x}|\vartheta)|_{\vartheta=0} \geqslant C, \tag{1.47}$$

où le seuil $C$ est défini par la probabilité donnée de fausse alarme

$$P\left\{\frac{\partial}{\partial\vartheta}\ln l(\mathbf{x}|\vartheta)\Big|_{\vartheta=0}\geqslant C|H_0\right\}=\alpha. \qquad (1.47')$$

Il est facile de généraliser la méthode localement optimale au critère de Bayes, en remplaçant la fonction de puissance par la valeur du risque conditionnel $r(\vartheta, \delta_n)$. Si $\delta_n^*$ est la règle de Bayes, la règle de Bayes localement optimale $\delta_n^{lo}$ sera donnée par la condition

$$r(0, \delta_n^{lo})=r(0, \delta_n^*),$$

$$\frac{\partial r(\vartheta, \delta_n^{lo})}{\partial\vartheta}\Big|_{\vartheta=0}=\frac{\partial r(\vartheta, \delta_n^*)}{\partial\vartheta}\Big|_{\vartheta=0} \qquad (1.48)$$

et la décision concernant la présence d'un signal est adoptée à partir de l'inégalité (1.47) où le seuil $C$ est défini par les valeurs données des pertes et par la probabilité a priori de présence du signal.

Cependant, on ne peut négliger les termes d'ordre plus élevé dans le développement du logarithme du rapport de vraisemblance dans le cas d'un signal faible que si l'échantillon est petit. Mais alors la probabilité de détection correcte, même avec des récepteurs optimaux, est du même ordre de grandeur que la probabilité de fausse alarme (voir fig. 1.1). Pour que la probabilité de détection correcte d'un signal faible soit voisine de l'unité, il y a lieu d'augmenter comme il faut la taille de l'échantillon. Il n'est plus possible de négliger les termes d'ordre élevé dans le développement du rapport de vraisemblance ni en général d'assurer la proximité voulue des caractéristiques des algorithmes localement optimal et optimal.

**1.2.8. Algorithmes asymptotiquement optimaux de détection des signaux.** Mieux adaptés aux problèmes pratiques sont les algorithmes asymptotiquement optimaux de détection des signaux noyés dans le bruit, dont les caractéristiques de travail tendent vers des caractéristiques rigoureusement optimales lorsque la taille de l'échantillon augmente indéfiniment et le rapport signal/bruit tend vers zéro *).

L'idée centrale de la théorie des algorithmes asymptotiquement optimaux de détection des signaux est dans la recherche d'une *statistique asymptotiquement suffisante* dont la distribution converge vers la loi *normale* lorsque la taille $n$ de l'échantillon augmente indéfiniment ($n\to\infty$) et l'amplitude du signal $\lambda\, s(t)$ tend vers zéro ($\lambda\to 0$). On a toute raison d'espérer que si l'on substitue cette sta-

*) Le terme « algorithme asymptotiquement optimal de détection (de discrimination) des signaux » sera entendu partout dans cet ouvrage seulement dans ce sens. Dans la littérature on attribue un sens plus large à l'optimum asymptotique. Par exemple, on considère comme asymptotiquement optimaux tous les algorithmes adaptatifs consistants.

tistique asymptotiquement suffisante dans la règle optimale du choix d'une décision donnée pour les distributions normales (1.1.9), on obtient un algorithme asymptotiquement optimal de détection des signaux noyés dans des bruits d'une large classe de distributions, algorithme qui possède une invariance partielle en ce sens que la distribution des bruits influera seulement sur le dispositif de formation d'une statistique suffisante. L'utilisation d'un algorithme asymptotiquement optimal dans le cas « sublimite » n'est justifiée que lorsque sa vitesse de convergence vers l'algorithme optimal est suffisamment grande. Ne cédant pas en qualité à des algorithmes construits à partir d'autres principes, les algorithmes asymptotiquement optimaux admettent une structure de systèmes de réalisation moins compliquée.

On peut s'attendre à ce que cette statistique normale et asymptotiquement suffisante existe pour les modèles de signaux et de bruits pour lesquels le rapport de vraisemblance admet une *factorisation* et donc le logarithme du rapport de vraisemblance est une *somme* de variables aléatoires.

Nous allons maintenant donner une définition rigoureuse de l'optimum asymptotique. Convenons de considérer comme optimal un algorithme de détection $\delta_n^0$ qui, pour une probabilité donnée de fausse alarme $\alpha_n$ et pour une taille donnée $n$ de l'échantillon, assure une probabilité minimale $\beta_n(\delta_n^0, \lambda_n)$ de perte du signal $\lambda_n s(t)$ (critère de Neyman-Pearson).

Soit $\lambda_1, \ldots, \lambda_n$ une suite d'amplitudes d'un signal convergeant d'une certaine manière vers zéro lorsque la taille $n$ de l'échantillon augmente indéfiniment. Considérons la suite $\{\delta_n\}$ d'algorithmes de détection des signaux d'amplitudes $\lambda_n$.

Désignons par $\beta_n(\delta_n, \lambda_n)$ la probabilité de perte du signal correspondant à $\delta_n$ et $\lambda_n$. Nous dirons que la suite des algorithmes $\{\delta_n^{ao}\}$ est *asymptotiquement optimale* *) si pour toute autre suite d'algorithmes $\{\delta_n\}$ on a la relation

$$\lim_{n\to\infty} [\beta_n(\delta_n, \lambda_n) - \beta_n(\delta_n^{ao}, \lambda_n)] \geqslant 0 \tag{1.49}$$

pour un niveau donné de fausse alarme $\alpha$. Ce niveau de fausse alarme doit également être pris en sens asymptotique, i.e.

$$\lim_{n\to\infty} \alpha_n(\delta_n) = \lim_{n\to\infty} \alpha_n(\delta_n^{ao}) = \alpha. \tag{1.49'}$$

La vitesse de convergence de la suite $\{\lambda_n\}$ vers zéro pour $n \to \infty$ n'est pas arbitraire. Il y a lieu de poser

$$\lambda_n = \gamma/\sqrt{n}, \tag{1.50}$$

*) Si $\beta_n(\delta_n, \lambda_n) - \beta_n(\delta_n^0, \lambda_n) \geqslant 0$ pour $n$ quelconque, l'algorithme $\delta_n^0$ est optimal suivant le critère de Neyman-Pearson pour la suite $\lambda_n$.

où $\gamma$ est une constante quelconque positive (bornée) proportionnelle au rapport signal/bruit. Une telle décroissance de l'amplitude du signal pour $n \to \infty$ signifie que le rapport signal/bruit est positif et borné (après le traitement du signal) pour un échantillon de taille quelconque. Lorsque l'amplitude du signal tend vers zéro plus rapidement, le rapport signal/bruit tend également vers zéro ; pour une convergence moins rapide de l'amplitude du signal vers zéro le rapport signal/bruit tend vers l'infini. Dans le premier cas la probabilité de détection correcte tend vers zéro, et dans le second vers l'unité. La restriction adoptée concernant le caractère des variations d'amplitude nous garantit contre la *singularité* *).

Ainsi, pour un algorithme asymptotiquement optimal sous la condition (1.50) il existe une limite différente de l'unité et de zéro

$$\lim_{n\to\infty} \beta_n (\delta_n^{\text{ao}}, \gamma/\sqrt{n}) = \beta(\delta^{\text{ao}}, \gamma), \tag{1.51}$$

$$0 < \beta(\delta^{\text{ao}}, \gamma) < 1. \tag{1.51'}$$

L'algorithme limite $\delta^{\text{ao}}$ utilise la statistique *normée* limite aux paramètres de distribution finis dépendant de $\gamma$.

Appuyons les propositions avancées par un exemple simple. On sait (voir tome II, page 41) que lors de la détection d'un signal constant positif $a$ noyé dans un bruit normal additif stationnaire indépendant de moyenne nulle et de variance $\sigma^2$, une comparaison de la somme normée

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{x_i}{\sigma}$$

d'observations indépendantes $x_i$ avec le seuil $C$ donne une probabilité minimale de perte du signal égale à **)

$$\beta = F(x_\alpha - \gamma), \tag{1.51''}$$

si le seuil $C$ est lié à la probabilité donnée $\alpha$ de fausse alarme par la relation

$$C = x_\alpha,$$

*) Lors de la détection du signal $\lambda s(t)$ noyé dans un bruit blanc de paramètre $d^2$ (un rapport signal/bruit à la sortie) dans (1.29) est égal à (voir (1.28))

$$d^2 = \frac{\lambda^2}{N_0} \int_0^T s^2(t)\,dt = \lambda^2 FT \frac{1}{N_0 F T_0} \int_0^T s^2(t)\,dt = \lambda^2 n\,(\text{s/b})_{\text{ent}} \sim \lambda^2 n = \gamma^2.$$

Pour $d^2 \to 0$, $\beta \to 1$ et pour $d^2 \to \infty$, $\beta \to 0$.

**) La statistique $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i/\sigma$ suit une loi normale de variance unité et de moyenne nulle pour l'hypothèse $H$ et égale à $\gamma$ pour l'alternative $K$.

où $x_\alpha$ est le quantile de la loi normale; $F(x)$ l'intégrale de Laplace et $\gamma = (a/\sigma)\sqrt{n}$. Ici $a/\sigma$ est le rapport signal/bruit à l'entrée du détecteur, la grandeur $\gamma$ est proportionnelle au rapport signal/bruit après traitement optimal de $n$ observations indépendantes. Si $0 < a/\sigma < \infty$, alors pour $n \to \infty$ la grandeur $\gamma$ croît indéfiniment et $\beta \to 0$ comme il se doit, vu la consistance de l'algorithme de détection. Si $n$ est fini et $a/\sigma \to 0$, on a $\gamma \to 0$ et par conséquent $x_{1-\beta} \to x_\alpha$, d'où $\beta \to 1 - \alpha$. Pour éviter les singularités relevées ($\beta = 0$ ou $\beta = 1 - \alpha$), il y a lieu de supposer que le rapport signal/bruit $a/\sigma$ à l'entrée décroît comme $\gamma/\sqrt{n}$, $0 < \gamma < \infty$. Dans ce cas l'équation

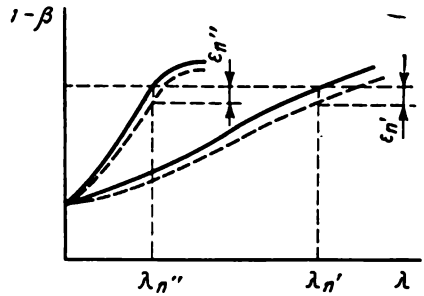



Fig. 1.2. Courbes de probabilité de détection correcte pour deux algorithmes: optimal suivant le critère de Neyman-Pearson (———) et asymptotiquement optimal (— — —).

(1.51″) est celle de la caractéristique de travail limite de l'algorithme optimal de détection d'un signal constant noyé dans un bruit normal additif.

Pour des échantillons de tailles importantes la caractéristique de l'algorithme asymptotiquement optimal de détection diffère peu de la caractéristique de l'algorithme optimal au sens du critère de Neyman-Pearson (fig. 1.2). Pour une probabilité de fausse alarme $\alpha$ donnée, les grandeurs $n'$, $\lambda_{n'}$ et $n''$, $\lambda_{n''}$ ($n'' > n'$, $\lambda_{n''} < \lambda_{n'}$) sont choisies de telle sorte que la probabilité $1 - \beta$ de détection correcte reste constante pour l'algorithme de Neyman-Pearson. Les quantités

$$|\beta_n(\delta_n^0, \gamma/\sqrt{n}) - \beta|,$$
$$|\alpha_n(\delta_n^0, \gamma/\sqrt{n}) - \alpha|$$

tendent vers zéro lorsque $\lambda_n \to 0$, $n \to \infty$, $\lambda_n \sqrt{n} \to \gamma > 0$.

Remarquons qu'en réalité l'amplitude du signal peut ne pas satisfaire à la condition (1.50). Mais cela n'empêche pas d'utiliser dans le cas « sublimite » un algorithme de détection asymptotiquement optimal dont la caractéristique de travail détermine la borne supérieure de la probabilité de détection correcte pour $\lambda \sim n^{-\nu}$, $\nu >$

$> \frac{1}{2}$ et la borne inférieure de cette probabilité pour $0 \leqslant \nu < \frac{1}{2}$ (c'est-à-dire lorsque l'algorithme de détection est consistant).

**1.2.9. Coefficient d'efficacité asymptotique relative.** Etablissons maintenant la mesure quantitative de l'optimum asymptotique. Comme nous l'avons noté dans 1.2.4, la qualité des algorithmes non paramétriques de détection est jugée d'après le coefficient d'efficacité relative (1.45) qui montre combien de fois le nombre d'observations que nécessite l'algorithme en question est plus élevé comparativement à un autre algorithme, pour des probabilités d'erreurs de première et de seconde espèce données.

Considérons de nouveau la classe $\Delta_\alpha$ d'algorithmes de détection consistants non paramétriques conservant constante la valeur $\alpha$ de la probabilité de fausse alarme. Soient $\delta_{n_k}$ et $\delta^*_{n^*_k}$ deux algorithmes de la classe $\Delta_\alpha$, et $\beta\,(w_{n_k}, \delta_{n_k})$, $\beta\,(w_{n_k}, \delta^*_{n^*_k})$ les probabilités de perte du signal pour une distribution donnée du signal et du bruit. Prenons ces probabilités égales à une probabilité de perte du signal donnée $\beta_0$:

$$\beta_{n_k}(w_{n_k}, \delta_{n_k}) = \beta_{n^*_k}(w_{n_k}, \delta^*_{n^*_k}) = \beta_0. \tag{1.52}$$

Pour $n_k \to \infty$ et $n^*_k \to \infty$ lorsque $k \to \infty$ on peut à partir de (1.52) trouver la limite *)

$$\rho_{\beta_0}(\delta, \delta^*) = \lim_{k\to\infty} (n_k/n^*_k) \tag{1.53}$$

que l'on appelle *coefficient d'efficacité asymptotique relative* de l'algorithme $\delta^*$ par rapport à l'algorithme $\delta$. Souvent cette limite ne dépend pas de $\alpha$ ni de $\beta_0$. Remarquons, cependant, que par suite de la consistance des algorithmes envisagés, les quantités $\beta\,(w_{n_k}, \delta_{n_k})$ et $\beta\,(w_{n^*_k}, \delta^*_{n^*_k})$ tendront vers zéro pour $k \to \infty$ pour une alternative $K$ donnée (distribution du signal et du bruit) différente de l'hypothèse $H$ (distribution du bruit). Pour que la valeur de la probabilité de perte du signal reste constante (égale à $\beta_0$) lorsque l'on passe à la limite, il faut que l'alternative et l'hypothèse se rapprochent. Ainsi, tout comme la notion d'optimum asymptotique, celle de critère asymptotique de qualité de détection d'un signal utilisant le coefficient d'efficacité asymptotique relative implique, afin d'éviter une singularité, que le rapport signal/bruit soit borné et différent de zéro lorsque la taille de l'échantillon augmente indéfiniment.

En vertu de la définition du coefficient d'efficacité asymptotique relative, pour trois algorithmes $\delta_1$, $\delta_2$, $\delta^*$ on a la relation

$$\rho_\beta(\delta_1, \delta^*)\,\rho_\beta(\delta^*, \delta_2) = \rho_\beta(\delta_1, \delta_2). \tag{1.53'}$$

*) $\delta$ et $\delta^*$ désignent ici des algorithmes limites (cf. (1.51)).

**1.2.10. Algorithmes de détection des signaux asymptotiquement les plus efficaces.** Soient $\delta_{n_k}$ et $\delta^*_{n^*_k}$ deux suites d'algorithmes de détection tels que les limites des probabilités de fausse alarme pour $n_k \to \infty$, $n^*_k \to \infty$, lorsque $k \to \infty$, coïncident:

$$\lim_{k\to\infty} \alpha_{n_k}(\delta_{n_k}) = \lim_{k\to\infty} \alpha_{n^*_k}(\delta^*_{n^*_k}) = \alpha, \tag{1.54}$$

et soit $\lambda_k$ une suite convergeant vers zéro pour $k \to \infty$ et telle que les limites des probabilités de perte du signal soient

$$\lim_{k\to\infty} \beta_{n_k}(\delta_{n_k}, \lambda_k) = \lim_{k\to\infty} \beta_{n^*_k}(\delta^*_{n^*_k}, \lambda_k) = \beta, \quad \alpha < \beta < 1. \tag{1.55}$$

Introduisons le coefficient d'efficacité asymptotique relative au niveau $\beta$ de la suite d'algorithmes $\delta^*_{n^*_k}$ par rapport à la suite $\delta_{n_k}$:

$$\rho_\beta(\delta, \delta^*) = \lim_{k\to\infty} (n_k/n^*_k). \tag{1.56}$$

Nous dirons que la suite d'algorithmes $\delta^*_{n^*_k}$ est *asymptotiquement la plus efficace* si pour $\delta_{n_k}$ quelconque et une valeur donnée de $\beta$ on a

$$\rho_\beta(\delta, \delta^*) \geqslant 1. \tag{1.57}$$

La suite $\delta^*_{n^*_k}$ est dite *uniformément asymptotiquement la plus efficace* si l'égalité (1.57) est vérifiée pour $\beta$ quelconque.

Toute règle *admissible* *) de choix d'une décision est asymptotiquement la plus efficace. On sait que, sous certaines restrictions, les règles admissibles appartiennent aux classes complètes **) de règles de Bayes et de règles basées sur le critère de Neyman-Pearson. Cependant, à la différence des règles optimales de Bayes et des règles optimales suivant le critère de Neyman-Pearson, les règles asymptotiquement les plus efficaces ne sont pas déterminées d'une manière univoque.

Soient $\delta_n$ et $\delta^*_n$ deux algorithmes réalisant une comparaison aux seuils des statistiques $y_n(\mathbf{x})$ et $y^*_n(\mathbf{x})$, et soit $\delta^*_n$ l'algorithme asymptotiquement le plus efficace. L'algorithme $\delta_n$ sera alors également asymptotiquement le plus efficace si seulement $|y_n - y^*_n| \to 0$ en probabilité pour $n \to \infty$.

---

*) Une règle $\delta^*$ de choix d'une décision est admissible s'il n'existe pas d'autre règle $\delta$ qui soit uniformément meilleure que $\delta^*$, c'est-à-dire qu'il n'existe pas de $\delta$ satisfaisant aux conditions $r(w, \delta) \leqslant r(w, \delta^*)$ (pour toutes les lois de répartition $w \in W$) $r(w, \delta) < r(w, \delta^*)$ (ne serait-ce que pour une seule $w \in W$), où $r(\cdot)$ est la fonction de risque.

**) Une classe C de règles de choix d'une décision est complète si pour tout $\delta \notin C$ on peut trouver un $\delta^* \in C$ uniformément meilleur que $\delta$, c'est-à-dire que dans une classe complète il y aura toujours une décision qui soit meilleure que toute décision n'appartenant pas à cette classe.

Cette particularité permet de rechercher les algorithmes asymptotiquement les plus efficaces doués des propriétés désirées, par exemple, ayant une structure plus simple que les algorithmes optimaux ou possédant certaines propriétés non paramétriques. Il est évident que les différents algorithmes asymptotiquement les plus efficaces peuvent différer par la vitesse de convergence des algorithmes « sublimite » vers les algorithmes limite et en ce sens un algorithme asymptotiquement le plus efficace sera meilleur que les autres.

**1.2.11. Calcul du coefficient d'efficacité asymptotique relative.** La définition (1.56) ne donne pas de méthode de calcul du coefficient d'efficacité asymptotique relative. Cependant, on peut montrer que si les limites finies existent (ce qui a lieu pour les algorithmes asymptotiquement optimaux)

$$\lim_{n\to\infty} \beta_n(\delta_n, \gamma/\sqrt{n}) = \beta(\delta, \gamma), \tag{1.58}$$

$$\lim_{n\to\infty} \beta_n(\delta_n^*, \gamma^*/\sqrt{n}) = \beta(\delta^*, \gamma^*), \tag{1.59}$$

le coefficient d'efficacité relative asymptotique de l'algorithme $\delta^*$ par rapport à $\delta$ est égal à

$$\rho_{\beta_0}(\delta, \delta^*) = (\gamma_0|\gamma_0^*)^2, \tag{1.60}$$

où $\gamma_0$ et $\gamma_0^*$ sont les racines minimales des équations

$$\beta(\delta, \gamma) = \beta_0, \quad \beta(\delta^*, \gamma^*) = \beta_0. \tag{1.61}$$

La formule (1.60) est intuitivement claire. En effet, pour $\beta_0$ donné, il existe des suites $n_k$ et $n_k^*$ croissantes indéfiniment lorsque $k \to \infty$, pour lesquelles $\lambda_k \to 0$ lorsque $k \to \infty$, et l'on a

$$\lim_{k\to\infty} \lambda_k \sqrt{n_k} = \gamma_0, \quad \lim_{k\to\infty} \lambda_k \sqrt{n_k^*} = \gamma_0^*, \tag{1.62}$$

d'où l'on déduit (1.60) à l'aide de (1.53).

En vertu de (1.49), (1.57) et (1.60), un algorithme asymptotiquement optimal est également asymptotiquement le plus efficace (fig. 1.3). Ceci permet d'utiliser le coefficient d'efficacité asymptotique relative comme une mesure quantitative de l'optimalité asymptotique.

**1.2.12. Coefficient d'efficacité asymptotique relative des algorithmes unilatéraux utilisant les statistiques asymptotiquement normales.** Nous allons envisager deux algorithmes consistants de détection d'un signal noyé dans des bruits: on prend la décision que

le signal est présent si

$$y_n^{(1)}(\mathrm{x}) \geqslant C_1 \quad (\text{algorithme } \delta_1), \tag{1.63}$$

$$y_n^{(2)}(\mathrm{x}) \geqslant C_2 \quad (\text{algorithme } \delta_2). \tag{1.63'}$$

Supposons que les statistiques $y_n^{(1)}(\mathrm{x})$ et $y_n^{(2)}(\mathrm{x})$ soient asymptotiquement normales *). Dans ce cas le coefficient d'efficacité relative asymptotique d'un algorithme par rapport à un autre peut être

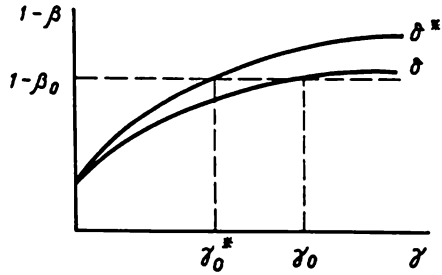



Fig. 1.3. Sur la définition du coefficient d'efficacité asymptotique relative.

présenté sous forme explicite par les caractéristiques probabilistes des processus observés.

Supposons comme précédemment que l'hypothèse $H$ signifie l'absence et l'alternative $K$ la présence du signal. Introduisons les désignations suivantes:

$$m_1\{y_n^{(i)}(\mathrm{x})\,|\,H\} = a_{nH}^{(i)}, \quad m_1\{y_n^{(i)}(\mathrm{x})\,|\,K\} = a_{nK}^{(i)}, \tag{1.64}$$

$$[M_2\{y_n^{(i)}(\mathrm{x})\,|\,H\}]^{1/2} = \sigma_{nH}^{(i)}, \quad [M_2\{y_n^{(i)}(\mathrm{x})\,|\,K\}]^{1/2} = \sigma_{nK}^{(i)}, \quad i = 1;\ 2, \tag{1.64'}$$

et supposons que

$$\lim_{n\to\infty} \frac{\sigma_{nH}^{(i)}}{\sigma_{nK}^{(i)}} = 1. \tag{1.65}$$

Si dans les algorithmes $\delta_1$ et $\delta_2$ les variables aléatoires $y_n^{(1)}(\mathrm{x})$ et $y_n^{(2)}(\mathrm{x})$ sont normales, pour une probabilité de fausse alarme donnée $\alpha$ les seuils $C_1$ et $C_2$ seront égaux (voir tome II, p. 44)

$$C_i = a_{nH}^{(i)} + \sigma_{nH}^{(i)} x_\alpha, \quad i = 1\ ;\ 2, \tag{1.66}$$

*) Dans les chapitres qui suivent, nous envisagerons des statistiques $y_n(\mathrm{x})$ qui seront asymptotiquement normales en vertu du théorème de la limite centrale. Comme on sait, ce théorème se formule pour des statistiques normées comme suit: $[y_n(\mathrm{x}) - m_1\{y_n(\mathrm{x})\}]/[M_2\{y_n(\mathrm{x})\}]^{1/2}$. Nous utiliserons souvent le terme « statistique asymptotiquement normale » pour des statistiques non normées, en nous basant sur le fait que la distribution d'une somme de grandeurs aléatoires indépendantes, pour un grand nombre de composantes, peut être approximée par une loi normale dont la moyenne et la variance dépendent du nombre de composantes.

et la probabilité de détection correcte $1 - \beta$ est liée à la probabilité de fausse alarme par la relation

$$x_{1-\beta} = x_\alpha \frac{\sigma_{nH}^{(i)}}{\sigma_{nK}^{(i)}} - \frac{a_{nK}^{(i)} - a_{nH}^{(i)}}{\sigma_{nK}^{(i)}}, \quad i = 1\ ;\ 2. \tag{1.67}$$

$x_\alpha$, $x_{1-\beta}$ sont ici les quantiles de la loi normale.

Pour $n \to \infty$, en vertu de (1.67) et compte tenu de (1.65) on obtient la caractéristique de travail limite

$$x_{1-\beta} = x_\alpha - \lambda_i, \tag{1.68}$$

où

$$\lambda_i = \lim_{n\to\infty} \frac{a_{nK}^{(i)} - a_{nH}^{(i)}}{\sigma_{nH}^{(i)}}. \tag{1.68'}$$

Si (comme c'est souvent le cas) la variance de la statistique $y_n$ (x) croît proportionnellement à la taille $n$ de l'échantillon, alors pour que la grandeur $\lambda_i$ soit finie et différente de zéro il faut que la différence $a_{nK}^{(i)} - a_{nH}^{(i)}$ augmente comme $\sqrt{n}$.

Supposons que la distribution du mélange signal — bruit est caractérisée par un paramètre positif $\vartheta$ (amplitude du signal), et qu'à la distribution du bruit correspond $\vartheta = 0$.

Introduisons les désignations

$$a_{nK}^{(i)} = a_{ni}(\vartheta), \quad a_{nH}^{(i)} = a_{ni}(0).$$

Pour des $\vartheta$ petits on a alors

$$a_{nK}^{(i)} - a_{nH}^{(i)} = a_{ni}(\vartheta) - a_{ni}(0) = \vartheta a'_{ni}(0) + o(\vartheta), \tag{1.69}$$

où

$$a'_{ni}(0) = \frac{d}{d\vartheta} a_{ni}(\vartheta)|_{\vartheta=0} \neq 0.$$

En posant $\gamma = \vartheta\sqrt{n}$, on obtient à partir de (1.68) et (1.69)

$$\lambda_i = \gamma\varepsilon_i, \tag{1.70}$$

où

$$\varepsilon_i = \lim_{n\to\infty} a'_{ni}(0)/[\sigma_{nH}^{(i)}\sqrt{n}]. \tag{1.70'}$$

On appelle souvent $\varepsilon_i$ mesure de l'efficacité de l'algorithme $\delta_i$. (1.70) et (1.68) conduisent à une équation linéaire par rapport à $\gamma$

$$x_{1-\beta} = x_\alpha - \gamma\varepsilon_i, \tag{1.71}$$

dont l'unique solution est

$$\gamma_i = \frac{x_\alpha - x_{1-\beta}}{\varepsilon_i}. \tag{1.71'}$$

Pour des probabilités d'erreurs données $\alpha$ et $\beta$, on déduit de (1.60) et (1.71') l'expression suivante pour le coefficient d'efficacité

relative de l'algorithme $\delta_1$ par rapport à l'algorithme $\delta_2$:

$$\rho = \left(\frac{\gamma_2}{\gamma_1}\right)^2 = \left(\frac{\varepsilon_1}{\varepsilon_2}\right)^2, \tag{1.72}$$

où $\varepsilon_1$ et $\varepsilon_2$ sont définis par (1.70').

Si les $r$ premières dérivées de $a_{ni}(\vartheta)$ par rapport à $\vartheta$ pour $\vartheta=0$ s'annulent, il y a lieu de substituer dans (1.70') la grandeur $a_{ni}^{(r+1)}$ (0).

**1.2.13. Coefficient d'efficacité asymptotique relative des algorithmes bilatéraux utilisant des statistiques asymptotiquement normales.** Considérons les algorithmes de détection bilatéraux consistants suivants: on prend la décision que le signal est présent si

$$|y_n^{(1)}(\mathbf{x})| \geqslant C_1 \quad \text{(algorithme } \delta_1\text{)}, \tag{1.73}$$

$$|y_n^{(2)}(\mathbf{x})| \geqslant C_2 \quad \text{(algorithme } \delta_2\text{)}. \tag{1.73'}$$

Conservant l'hypothèse que les statistiques $y_n^{(1)}(\mathbf{x})$ et $y_n^{(2)}(\mathbf{x})$ sont asymptotiquement normales et les désignations (1.64) et (1.64') nous obtenons pour $\alpha$ et $\beta$ donnés les expressions suivantes des caractéristiques de travail des algorithmes bilatéraux (voir tome II, (1.132))

$$\beta = F(x_{\alpha/2} - \gamma_1\varepsilon_1) - F(-x_{\alpha/2} - \gamma_1\varepsilon_1), \tag{1.74}$$

$$\beta = F(x_{\alpha/2} - \gamma_2\varepsilon_2) - F(-x_{\alpha/2} - \gamma_2\varepsilon_2). \tag{1.74'}$$

En égalant les seconds membres des égalités (1.74) et (1.74') et compte tenu de la monotonie des fonctions égalées, nous nous convainquons que la formule (1.72) reste vraie pour des algorithmes de détection bilatéraux.

**1.2.14. Mesure quantitative de stabilité de l'algorithme asymptotiquement optimal de détection des signaux.** En vue des applications pratiques des algorithmes asymptotiquement optimaux de détection des signaux, il est tout naturel de se poser la question de savoir combien les caractéristiques de détection sont sensibles aux variations de la distribution des probabilités des bruits par rapport à la distribution pour laquelle l'algorithme utilisé est asymptotiquement optimal. Nous appellerons *stabilité* la propriété d'un détecteur de conserver ses caractéristiques dans certaines limites de variation des bruits.

Soit $w$ la distribution d'un bruit par rapport auquel l'algorithme $\delta^*(w)$ réalisé dans le détecteur est asymptotiquement optimal, et soit $u$ la distribution du bruit lorsque les conditions ont changé. Nous allons adopter pour mesure de stabilité le coefficient $\rho$ d'efficacité relative asymptotique de l'algorithme $\delta_u^*(w)$, utilisé pour un bruit « étranger » de distribution $u$, par rapport à l'algorithme $\delta^*(u)$ asymptotiquement optimal pour ce bruit « étranger ».

Supposons que les algorithmes $\delta_u^*(w)$ et $\delta^*(u)$ réalisent une comparaison avec un seuil de statistiques asymptotiquement norma-

les et que les paramètres (moyenne et variance) de la distribution limite normale pour l'hypothèse $H$ (il n'y a pas de signal) et pour l'alternative $K$ (le signal est présent) sont

| algorithme | $\delta_u^*(w)$ | $\delta^*(u)$ |
|---|---|---|
| hypothèse $H$ | $0,\ \sigma_1^2$ | $0,\ \sigma^2$ |
| alternative $K$ | $\gamma a_1,\ \sigma_1^2$ | $\gamma a,\ \sigma^2$. |

Pour une probabilité donnée de fausse alarme $\alpha$, les probabilités de perte du signal sont égales à

$$\beta_1(\delta_u^*,\ \gamma) = F\left(x_\alpha - \frac{\gamma a_1}{\sigma_1}\right) \tag{1.75}$$

lorsque l'on utilise l'algorithme $\delta_u^*(w)$, et à

$$\beta_1(\delta^*,\ \gamma) = F\left(x_\alpha - \frac{\gamma a}{\sigma}\right) \tag{1.75'}$$

lorsque l'on utilise l'algorithme $\delta^*(u)$, $x_\alpha$ étant le quantile de la loi normale. En substituant (1.75) et (1.75') dans (1.61), on trouve $\gamma_0$ et $\gamma_0^*$

$$\gamma_0 = (x_\alpha - x_{1-\beta})\frac{\sigma}{a},\quad \gamma_0^* = (x_\alpha - x_{1-\beta})\frac{\sigma_1}{a_1}, \tag{1.76}$$

qui sont les racines uniques des équations (1.61) en vertu de la monotonie de l'intégrale de Laplace $F(z)$.

En substituant (1.76) dans (1.60) on trouve le coefficient d'efficacité relative asymptotique de l'algorithme $\delta_u^*(w)$ par rapport à $\delta^*(u)$

$$\rho = (a_1/\sigma_1)^2\,(\sigma/a)^2. \tag{1.77}$$

Remarquons que les rapports $a/\sigma$ et $a_1/\sigma_1$ peuvent être considérés comme les mesures des « distances » entre les distributions limites pour l'hypothèse (il n'y a pas de signal) et l'alternative (le signal est présent) ou comme le rapport « signal/bruit ».

**1.2.15. Résumé.** La classification des problèmes de la synthèse dans les conditions d'indétermination a priori que nous avons esquissée dans ce paragraphe pour différents critères de qualité permet de distinguer les méthodes essentielles suivantes utilisées pour surmonter l'indétermination a priori : 1) les méthodes de la statistique non paramétrique, 2) les méthodes asymptotiques, 3) les méthodes d'adaptation utilisant des échantillons d'apprentissage, 4) les méthodes de la statistique paramétrique s'appuyant sur les principes d'invariance, de similitude et du minimax.

Nous passons maintenant à une caractéristique détaillée de ces méthodes. Le choix des matières exposées, comprenant des exemples d'application pratique, reflète le côté radiotechnique spécifique des problèmes de la détection et de la séparation des signaux noyés dans des bruits.

La formalisation des résultats généraux permet de les utiliser dans d'autres domaines de la science et de la technique.

CHAPITRE 2

# MÉTHODES NON PARAMÉTRIQUES DE SYNTHÈSE STATISTIQUE

## 2.1. CRITÈRES DE CONFORMITÉ

**2.1.1. Vérification de l'hypothèse sur la forme de la fonction de répartition.** Supposons que l'on ait un échantillon $\mathbf{x} = (x_1, \ldots, x_n)$ d'observations indépendantes appartenant à une distribution inconnue $G_1(x)$. Il y a lieu de tester l'hypothèse $H$ suivant laquelle $G_1(x)$ coïncide avec une fonction de répartition continue $F_1(x)$ donnée (hypothétique), contre l'alternative $K$ que $G_1(x) \neq F_1(x)$. Parfois on envisage les alternatives unilatérales $K^+$: $F_1(x) > G_1(x)$ ou $K^-$: $F_1(x) < G_1(x)$.

La solution de ces problèmes est basée sur l'utilisation d'une fonction de répartition empirique (voir tome II, 2.1.1)

$$F_1^*(x, \mathbf{x}^{(\ )}) = \frac{1}{n} \sum_{k=1}^{n} u(x - x^{(k)}), \tag{2.1}$$

où $\mathbf{x}^{(\ )} = (x^{(1)}, \ldots, x^{(n)})$ est un échantillon ordonné (vecteur des statistiques d'ordre) et $u(z)$ la fonction du saut unitaire égale à

$$u(z) = \begin{cases} 0, & z < 0, \\ 1, & z \geqslant 0. \end{cases} \tag{2.1'}$$

Si l'hypothèse $H$ est vraie, la fonction de répartition empirique converge en probabilité lors d'une augmentation illimitée de la taille $n$ de l'échantillon vers la fonction hypothétique, i.e. pour $n \to \infty$ la suite des variables aléatoires

$$\zeta_n = \sup_x |F_1(x) - F_1^*(x, \mathbf{x}^{(\ )})| \tag{2.2}$$

converge en probabilité vers zéro.

Pour un échantillon de taille finie les fonctions de répartition empirique et hypothétique différerons l'une de l'autre. Introduisons la mesure numérique de la distinction (*critère de conformité*) $d_n(F_1^*, F_1)$ des fonctions de répartition empirique $F_1^*(x, \mathbf{x}^{(\ )})$ et hypothétique $F_1(x)$. Dans nos décisions nous allons nous conformer à la règle suivante: si la statistique $d_n(F_1^*, F_1)$ dépasse un certain seuil, l'hypothèse $H$ est rejetée. La valeur de ce seuil est trouvée comme

d'habitude pour une probabilité donnée de faux rejet de l'hypothèse, c'est-à-dire à partir de la relation

$$P\{d_n(F_1^*, F_1) > d_\alpha | H\} = \alpha. \tag{2.3}$$

Si la distribution de la statistique $d_n(F_1^*, F_1)$ *ne dépend pas de l'hypothèse* $H$, i.e. de la forme de la distribution $F_1(x)$, la règle de rejet de l'hypothèse pour

$$d_n(F_1^*, F_1) > d_\alpha \tag{2.4}$$

est *non paramétrique*. Conformément à la terminologie de 1.2.4, l'algorithme (2.4) appartient à la classe $\Delta_\alpha$ des algorithmes de détection consistants non paramétriques si $F_1(x)$ est une famille de distributions du bruit non paramétrique.

Comme dans le cas de la vérification des hypothèses concernant les paramètres d'une distribution de forme donnée, l'hypothèse $H$ peut être *simple* ou *composite*. L'hypothèse $H$: $G_1(x) = F_1(x)$ est simple si la fonction $F_1(x)$ est entièrement déterminée. Par exemple, avancer que les observations appartiennent à une distribution normale à valeur moyenne et variance connues c'est faire une hypothèse simple. Une hypothèse $H$: $G_1(x) = F_1(x; \vartheta)$ est composite si la fonction $F_1$ est donnée aux paramètres inconnus $\vartheta = (\vartheta_1, \ldots, \vartheta_m)$, $\vartheta \in \Theta$, $m > 1$, près. Par exemple, l'hypothèse sur la normalité des observations lorsque la moyenne et la variance sont inconnues est une hypothèse composite.

**2.1.2. Critères de conformité.** Considérons la classe des critères de conformité dépendant de la différence des fonctions de répartition empirique $F_1^*$ et hypothétique $F_1$

$$d_n(F_1^*, F_1) = \psi\{F_1^*(x, \mathbf{x}^{(n)}) - F_1(x)\}. \tag{2.5}$$

Si $F_1(x)$ est continue, pour l'hypothèse $H$ la fonction de répartition de la statistique (2.5) *ne dépend pas de* $F_1(x)$, i.e. le critère de conformité est dans ce cas non paramétrique.

Pour démontrer que le critère (2.5) est non paramétrique nous avons besoin du résultat suivant de la théorie des probabilités. Supposons que les variables aléatoires $\eta$ et $\xi$ sont liées par la relation fonctionnelle $\eta = F_1(\xi)$. Supposons de plus que $F_1(x)$ est continue et représente la fonction de répartition de la variable aléatoire $\xi$. La variable aléatoire $\eta$ est alors répartie *uniformément* sur l'intervalle (0,1). Si $x_1, \ldots, x_n$ est un échantillon appartenant à la distribution continue $F_1(x)$, les variables aléatoires $y_k = F_1(x_k)$, $k = 1, \ldots, n$, forment un échantillon de la distribution uniforme

$$\Phi_1(y) = y, \ 0 \leqslant y \leqslant 1. \tag{2.6}$$

La fonction de répartition empirique

$$\Phi_1^*(y, \mathrm{y}^{(\ )}) = \frac{k}{n}, \quad y^{(k)} < y \leqslant y^{k+1}, \tag{2.7}$$

où $\mathrm{y}^{(\ )} = (y^{(1)}, \ldots, y^{(n)})$ est le vecteur des statistiques d'ordre de l'échantillon $\mathrm{y} = (y_1, \ldots, y_n)$.

En vertu de (2.6) et (2.7) pour $0 \leqslant y \leqslant 1$

$$\Phi_1(y) - \Phi_1^*(y, \mathrm{y}^{(\ )}) = y - \Phi_1^*(y, \mathrm{y}^{(\ )}) =$$

$$= F_1(x) - \Phi_1^*[F_1(x), F_1(\mathrm{x}^{(\ )})] = F_1(x) - F_1^*(x, \mathrm{x}^{(\ )}).$$

Par conséquent, pour l'hypothèse $H$ selon laquelle l'échantillon appartient à la distribution continue $F_1(x)$, la distribution de la fonctionnelle de la différence entre la fonction de répartition empirique et celle hypothétique coïncide avec la distribution de la fonctionnelle de la différence mentionnée pour le cas où l'échantillon appartient à une distribution uniforme sur l'intervalle (0,1). Autrement dit, la distribution de la fonctionnelle $\psi\{F_1^*(x, \mathrm{x}^{(\ )}) - F_1(x)\}$ ne dépend pas de la forme de $F_1(x)$.

**2.1.3. Critères de conformité de Kolmogorov et de Smirnov.** En vertu du critère de Kolmogorov, la mesure de la différence des distributions empirique et hypothétique est donnée par la statistique (voir (2.2))

$$d_n = \sup_x |F_1(x) - F_1^*(x, \mathrm{x}^{(\ )})|, \tag{2.8}$$

i.e. la borne supérieure, sur tous les $x$, des modules de la différence des distributions mentionnées.

Une propriété importante du critère de Kolmogorov est qu'il est non paramétrique, c'est-à-dire que si la fonction de répartition $F_1$ est continue, la fonction de répartition de la statistique (2.5) *ne dépend pas* de $F_1(x)$. Cette propriété découle directement du résultat formulé dans 2.1.2 car

$$d_n = \sup_x |F_1(x) - F_1^*(x, \mathrm{x}^{(\ )})| = \sup_y |\Phi_1(y) - \Phi_1^*(y, \mathrm{y}^{(\ )})|,$$

c'est-à-dire que la distribution de la statistique (2.8) coïncide avec celle de la statistique $\sup_y |\Phi_1(y) - \Phi_1^*(y, \mathrm{y}^{(\ )})|$ qui ne dépend pas de la distribution $F_1(x)$.

L'expression de la fonction de répartition de la statistique (2.8) pour un échantillon de taille finie est très compliquée (voir par exemple [4.7]). A. Kolmogorov a trouvé une expression asymptotique suffisamment simple pour cette fonction. Si l'hypothèse $H$ est vraie alors, pour $n \to \infty$ et sous l'hypothèse supplémentaire de

continuité de $F_1(x)$ la fonction de répartition *) de la grandeur $\sqrt{n}\, d_n$ s'approche asymptotiquement de

$$P\{\sqrt{n}\, d_n \leqslant z\} \sim \mathrm{k}(z) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2z^2}, \quad z > 0. \tag{2.9}$$

La fonction k $(z)$ peut s'écrire comme suit:

$$\mathrm{k}(z) = 1 - 2 \sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2z^2}, \quad z > 0. \tag{2.10}$$

D'une autre part, en écrivant k $(z)$ sous la forme

$$\mathrm{k}(z) = \sum_{k=-\infty}^{\infty} e^{-2(2k)^2z^2} - \sum_{k=-\infty}^{\infty} e^{-2(2k+1)^2z^2}$$

et en utilisant la formule de sommation de Poisson

$$\sum_{k=-\infty}^{\infty} h(k) = \sum_{k=-\infty}^{\infty} \int_{-\infty}^{\infty} h(x)\, e^{2\pi ikx}\, dx,$$

on peut écrire cette fonction comme suit:

$$\mathrm{k}(z) = \frac{\sqrt{2\pi}}{z} \sum_{k=1}^{\infty} \exp\left\{-\frac{\pi^2(2k-1)^2}{8z^2}\right\}. \tag{2.11}$$

Il est commode d'utiliser (2.10) pour $z > 1$ et (2.11) pour $z \ll 1$.

Si $\alpha$ est la probabilité donnée de rejet injuste de l'hypothèse (niveau significatif) on obtient à partir de (2.10)

$$P\{d_n > d_\alpha\} = P\{d_n\sqrt{n} > d_\alpha\sqrt{n}\} \sim 1 - \mathrm{k}(d_\alpha\sqrt{n}) =$$

$$= 2 \sum_{k=1}^{\infty} (-1)^{k-1} \exp\{-2k^2 d_\alpha^2 n\} = \alpha \tag{2.12}$$

ou sous une autre forme, en utilisant (2.11), on a

$$P\{d_n > d_\alpha\} \sim \frac{1}{d_\alpha}\sqrt{\frac{2\pi}{n}} \sum_{k=1}^{\infty} \exp\left\{-\frac{\pi^2(k-1)^2}{8nd_\alpha^2}\right\} = \alpha. \tag{2.13}$$

La série dans (2.12) converge rapidement et souvent en première approximation on peut se limiter au premier terme

$$2e^{-2nd_\alpha^2} = \alpha \tag{2.14}$$

*) La fonction de répartition envisagée peut s'écrire comme suit [3]:

$$P\{\sqrt{n}\, d_n \leqslant z\} = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2z^2}\left[1 - \frac{2k^2x}{3\sqrt{n}} + o\left(\frac{1}{n}\right)\right].$$

ou

$$d_\alpha = \left(\frac{1}{2n} \ln \frac{2}{\alpha}\right)^{1/2}. \tag{2.15}$$

La règle de vérification de l'hypothèse $H$ contre l'alternative bilatérale $K$ est la suivante : si, pour l'échantillon observé et pour une valeur donnée de $\alpha$, on a

$$d_n = \sup_x |F_1(x) - F_1^*(x, \mathrm{x}^{(\,)})| > d_\alpha, \tag{2.16}$$

l'hypothèse selon laquelle l'échantillon a été extrait de la distribution $F_1(x)$ *est rejetée*. Pour des tailles importantes des échantillons ($n > 50$) le seuil $d_\alpha$ est donné par l'une des formules (2.12) à (2.15). Si l'on construit deux courbes en escalier $y = F_1^*(x, \mathrm{x}^{(\,)}) + d_\alpha$ et $y = F_1^*(x, \mathrm{x}^{(\,)}) - d_\alpha$, on peut affirmer avec un certain degré de certitude déterminé par le niveau significatif donné que la fonction de répartition hypothétique $F_1(x)$ se trouve à l'intérieur de la bande limitée par les courbes mentionnées (limites de confiance).

On trouvera dans [4] une table des valeurs des seuils $d_\alpha$ pour $\alpha = 0{,}01$ et $\alpha = 0{,}05$, pour les valeurs de $n$ allant de 5 à 100, les valeurs exactes étant calculées à partir de la distribution $d_n$ pour des $n$ finis, et les valeurs asymptotiques à l'aide de la fonction k ($z$). Cette table montre que pour $\alpha \geqslant 0{,}01$ et $n > 50$ l'erreur relative de la valeur asymptotique par rapport à la valeur exacte ne dépasse pas 2 %. On peut trouver des tables plus complètes dans [6].

Pour l'alternative unilatérale $K^+$ : $F_1(x) > F_1^*(x, \mathrm{x}^{(\,)})$ on utilise le critère unilatéral de Smirnov

$$d_n^+ = \sup_x [F_1(x) - F_1^*(x, \mathrm{x}^{(\,)})], \tag{2.17}$$

et la règle de vérification de l'hypothèse $H$ contre $K^+$ prescrit à rejeter l'hypothèse, pour un niveau significatif donné $\alpha'$, si $d_n^+ > d_{\alpha'}$, de plus

$$\lim_{n\to\infty} P\{\sqrt{n}\, d_n^+ \geqslant z\} = e^{-2z^2} \tag{2.18}$$

et par conséquent pour $n \to \infty$

$$\alpha' \sim e^{-2nd_{\alpha'}^2}. \tag{2.18'}$$

Les limites de confiance sont données par les courbes en escalier $y = F_1^*(x, \mathrm{x}^{(\,)})$, $y = F_1^*(x, \mathrm{x}^{(\,)}) + d_{\alpha'}$.

Pour l'alternative unilatérale $K^-$ : $F_1(x) < F_1(x, \mathrm{x}^{(\,)})$ on utilise le critère unilatéral suivant

$$d_n^- = \inf_x [F_1(x) - F_1^*(x, \mathrm{x}^{(\,)})]. \tag{2.19}$$

Il évident que $P\{d_n^+ > d_{\alpha'}\} = \alpha'$ entraîne $P\{d_n^- < -d_{\alpha'}\} = \alpha'$ et

$$\alpha = P\{d_n > d_\alpha\} = P\{(d_n^+ > d_{\alpha'}) \cup (d_n^- < -d_{\alpha'})\} \leqslant 2\alpha'.$$

Dans certains cas la fonction de répartition hypothétique peut dépendre du paramètre inconnu $\vartheta$ (scalaire ou vectoriel). On vérifie alors l'hypothèse composite $H$ selon laquelle l'échantillon observé appartient à la distribution $F_1(x; \vartheta)$. Dans le cas envisagé au lieu de (2.5) on peut utiliser la statistique

$$\hat{d}_n = \sup_x |F_1^*(x, \mathbf{x}^{(\ )}) - F_1(x, \hat{\vartheta})|, \tag{2.20}$$

où $\hat{\vartheta}$ est l'estimation consistante du paramètre $\vartheta$ (par exemple, l'estimation du maximum de vraisemblance; pour plus de détails voir [28]). Cependant la statistique $\hat{d}_n$ n'est pas en général non paramétrique, c'est-à-dire que sa distibution dépend de $F_1$.

**2.1.4. Critère de conformité de Renyi.** A la différence des critères de Kolmogorov et de Smirnov qui sont des fonctionnelles de la différence des distributions hypothétique et empirique, A. Renyi a introduit des critères de conformité sous la forme de fonctionnelles de l'*écart relatif* des distributions empirique et hypothétique; le critère bilatéral

$$d_n = \sup_x \left| \frac{F_1^*(x, \mathbf{x}^{(\ )}) - F_1(x)}{F_1(x)} \right|, \tag{2.21}$$

$$0 < a \leqslant F_1(x) \leqslant b \leqslant 1,$$

et le critère unilatéral

$$d_n^+ = \sup_x \frac{F_1^*(x, \mathbf{x}^{(\ )}) - F_1(x)}{F_1(x)}, \tag{2.21'}$$

$$0 < a \leqslant F_1(x) \leqslant b \leqslant 1.$$

Les critères de Renyi sont non paramétriques par rapport à l'hypothèse $H$, car ce sont des fonctionnelles de la différence des fonctions de répartition empirique et hypothétique.

Pour les statistiques (2.21) et (2.21') on a obtenu des distributions asymptotiques lorsque $n \to \infty$. Ces distributions (pour $b = 1$)

sont [35]

$$\lim_{n\to\infty} P\{\sqrt{n}d_n \leqslant z\} = \frac{4}{\pi}\sum_{k=0}^{\infty}(-1)^k \frac{1}{2k+1} \times$$

$$\times \exp\left\{-\frac{(2k+1)^2\pi^2(1-a)}{8z^2a}\right\}, \quad z>0,\ a>0, \tag{2.22}$$

$$\lim_{n\to\infty} P\{\sqrt{n}d_n^+ \leqslant z\} = \sqrt{\frac{2}{\pi}} \int_0^{z\sqrt{\frac{a}{1-a}}} e^{-t^2/2}\,dt =$$

$$= 2F\left(z\sqrt{\frac{a}{a-1}}\right) - 1, \quad z>0,\ a>0, \tag{2.22'}$$

où $F(u)$ est l'intégrale de Laplace.

On trouvera dans [3, 35] les tables de valeurs de la distribution asymptotique (2.22).

**2.1.5. Critère de conformité de Mises.** En vertu de ce critère la mesure quantitative de conformité est pour une taille donnée de l'échantillon la valeur moyenne de l'écart quadratique pondéré de la distribution empirique par rapport à la distribution hypothétique, soit

$$d_n = \int_{-\infty}^{\infty} [F_1^*(x, \mathrm{x}^{(\ )}) - F_1(x)]^2 \psi[F_1(x)]\,dF_1(x), \tag{2.23}$$

où $\psi(u)$ est la fonction de pondération. Pour $\psi \equiv 1$, on a

$$d_n = m_1\{[F_1^*(x, \mathrm{x}^{(\ )}) - F_1(x)]^2\} = \int_{-\infty}^{\infty} [F_1^*(x, \mathrm{x}^{(\ )}) - F_1(x)]^2\,dF_1(x). \tag{2.24}$$

En substituant dans (2.24) l'expression de $F_1^*(x, \mathrm{x}^{(\ )})$ donnée par (2.1) et en intégrant on obtient

$$d_n = \frac{1}{12n^2} + \frac{1}{n}\sum_{k=1}^{n}\left[F_1(x^{(k)}) - \frac{2k-1}{2n}\right]^2. \tag{2.25}$$

Il est facile de voir [10] que

$$m_1\{d_n\} = \frac{1}{6n}, \quad M_2\{d_n\} = \frac{4n-3}{180n^3}. \tag{2.26}$$

Le critère de Mises est non paramétrique par rapport à l'hypothèse $H$, car c'est une fonctionnelle de la différence des fonctions de répar-

tition empirique et hypothétique. L'expression exacte de la distribution de la grandeur $\sqrt{n}d_n$ pour le critère de Mises, tout comme pour les critères envisagés ci-dessus, est très compliquée, mais pour $n > 40$ elle s'approche de la distribution limite. Donnons à titre d'exemple quelques quantiles de cette distribution limite ($n \to \infty$):

| | | | | | |
|---|---|---|---|---|---|
| $\alpha$ | 0,5 | 0,4 | 0,3 | 0,2 | 0,1 |
| $d_\alpha$ | 0,1184 | 0,1467 | 0,1843 | 0,2412 | 0,3673 |
| $\alpha$ | 0,05 | 0,03 | 0,02 | 0,01 | 0,001 |
| $d_\alpha$ | 0,4614 | 0,5489 | 0,6198 | 0,7435 | 1,1679 |

**2.1.6. Critère de conformité du $\chi^2$.** Ce critère est l'un des plus usités dans les applications pratiques.

Divisons le domaine de définition de la fonction de répartition hypothétique $F_1(x)$ en un nombre fini d'intervalles disjoints $\Delta_i$, $i = 1, \ldots, l$. Désignons par $p_i$, $\sum_{i=1}^{l} p_i = 1$ la probabilité a priori pour une valeur échantillonnée de se trouver dans l'intervalle $\Delta_i$. Soit $\nu_i$ le nombre de valeurs de l'échantillon $\mathrm{x} = (x_1, \ldots, x_n)$ qui tombent dans l'intervalle $\Delta_i$. Il est évident que $\sum_{i=1}^{l} \nu_i = n$.

Prenons pour mesure d'écart entre les valeurs observées et les données a priori la grandeur

$$d_n = \sum_{i=1}^{l} \frac{1}{np_i} (\nu_i - np_i)^2. \tag{2.27}$$

Pearson (voir [11], § 30.1) a montré que si l'hypothèse testée $H$ selon laquelle l'échantillon x appartient à la distribution $F_1(x)$ est vraie, pour $n \to \infty$ la distribution de la statistique (2.27) tend asymptotiquement vers une distribution du $\chi^2$ à $l - 1$ degrés de liberté qui ne dépend pas de la distribution hypothétique $F_1(x)$.

Soit $\chi_\alpha$ l'$\alpha$-quantile d'une variable aléatoire répartie suivant la loi du $\chi^2$ à $l - 1$ degrés de liberté, c'est-à-dire $P\{\chi^2 > \chi_\alpha^2\} = \alpha$. La règle de vérification de l'hypothèse $H$ se formule alors comme suit: l'hypothèse est rejetée si $d_n > \chi_\alpha^2$. La probabilité de la rejeter à tort est égale à $\alpha$.

Outre le défaut commun des critères de conformité d'avoir une distribution de la statistique (2.27) très compliquée pour un échantillon de taille finie, le critère du $\chi^2$ présente l'inconvénient lié à la division arbitraire de l'axe réel en intervalles $\Delta_i$, sans rapport

à la forme de la fonction $F_1(x)$ ni au groupement des valeurs échantillonnées *).

Les quantiles de la répartition du $\chi^2$ sont indiqués dans de nombreux ouvrages et tables spéciales (voir, par exemple, tome II, annexe VIII, ainsi que [3]).

**2.1.7. Vérification de l'hypothèse d'appartenance de deux échantillons à une même distribution.** Soient $\mathbf{x} = (x_1, \ldots, x_{n_1})$, $\mathbf{y} = (y_1, \ldots, y_{n_2})$ des échantillons à éléments indépendants appartenant aux distributions $F_{1x}(z)$ et $F_{1y}(z)$ respectivement. Pour de nombreuses applications, il est utile de savoir vérifier l'hypothèse $H: F_{1x}(z) = F_{1y}(z)$, contre l'alternative $K: F_{1x}(z) \neq F_{1y}(z)$.

Si $F_{1x}(z)$ est une distribution du bruit, et $F_{1y}(z)$ une distribution du mélange du signal et du bruit, le problème formulé est un problème de la détection d'un signal dans l'échantillon observé y lorsque l'on dispose de l'échantillon *d'apprentissage* du bruit x. Ce problème apparaît chaque fois où il y a lieu de vérifier l'*homogénéité* d'un échantillon composé, obtenu à partir de plusieurs expériences ou essais.

Il est clair qu'avec l'hypothèse $H$ et une fonction de répartition $F_1(z)$ commune des échantillons x et y continue, la transformation $u = F_1(z)$ fait correspondre à l'échantillon (x, y) un échantillon homogène issu d'une distribution uniforme sur l'intervalle (0,1). Par conséquent, les critères qui sont des fonctionnelles de la différence des fonctions de répartition empiriques $F_{1x}^*(z, \mathbf{x}^{(\cdot)}) - F_{1x}^*(z, \mathbf{y}^{(\cdot)})$ sont *non paramétriques* relativement à l'hypothèse $H$ (voir 2.1.2).

On peut utiliser pour la vérification des hypothèses de l'appartenance de deux échantillons à une même distribution les trois critères suivants, analogues aux critères envisagés ci-dessus:
le critère de Kolmogorov

$$d_{n_1 n_2} = \sup_z |F_{1x}^*(z, \mathbf{x}^{(\cdot)}) - F_{1y}^*(z, \mathbf{y}^{(\cdot)})|, \tag{2.28}$$

le critère de Smirnov

$$d_{n_1 n_2}^+ = \sup_z \{F_{1x}^*(z, \mathbf{x}^{(\cdot)}) - F_{1y}^*(z, \mathbf{y}^{(\cdot)})\} \tag{2.29}$$

et le critère de Mises

$$d_{n_1 n_2}^{\mathrm{M}} = \int_{-\infty}^{\infty} [F_{1x}^*(z, \mathbf{x}^{(\cdot)}) - F_{1y}^*(z, \mathbf{y}^{(\cdot)})]^2 \, dF^*(z, \mathbf{z}^{(\cdot)}, \mathbf{y}^{(\cdot)}), \tag{2.30}$$

*) Remarquons que le critère du $\chi^2$ convient également pour les distributions discrètes. Si $x_1, \ldots, x_l$ sont les valeurs possibles de la variable aléatoire $\xi$, et $p_i$ la probabilité a priori pour que $\xi = x_i$, alors dans la formule (2.27) $\nu_i$ est le nombre de valeurs échantillonnées (dans l'échantillon de taille $n$ de cette distribution) égales à $x_i$.

où

$$F^*(z, x^{(\,)}, y^{(\,)}) = \frac{n_1 F^*_{1x}(z, x^{(\,)}) + n_2 F^*_{1y}(z, y^{(\,)})}{n_1 + n_2}. \tag{2.31}$$

Tous ces critères sont non paramétriques par rapport à l'hypothèse $H$.

Comme l'a montré N. Smirnov (voir [6], § 63), pour l'hypothèse $H$ on a

$$\lim_{\substack{n_1 \to \infty \\ n_2 \to \infty}} P\{\sqrt{N}\, d_{n_1 n_2} > z\} = 1 - \mathrm{k}(z), \quad z > 0, \tag{2.32}$$

$$\lim_{\substack{n_1 \to \infty \\ n_2 \to \infty}} P\{\sqrt{N}\, d^+_{n_1 n_2} > z\} = e^{-2z^2}, \quad z > 0, \tag{2.33}$$

où $1/N = 1/n_1 + 1/n_2$ et $\mathrm{k}(z)$ est la fonction définie par (2.9).

On peut maintenant formuler la règle de vérification de l'hypothèse $H$ à l'aide des critères de Kolmogorov-Smirnov. Pour l'alternative $F_{1x}(z) \neq F_{1y}(z)$, l'hypothèse $H$ est rejetée si

$$d_{n_1 n_2} > d_\alpha, \tag{2.34}$$

où le seuil $d_\alpha$ est défini pour une probabilité donnée de l'erreur (niveau significatif) à partir de la condition (pour $N \gg 1$)

$$1 - \mathrm{k}(\sqrt{N} d_\alpha) = \alpha. \tag{2.34'}$$

On peut utiliser pour première approximation de (2.34') la formule suivante:

$$d_\alpha = \left(\frac{1}{2N} \ln \frac{2}{\alpha}\right)^{\frac{1}{2}}. \tag{2.34''}$$

Pour l'alternative unilatérale $F_{1x}(z) > F_{1y}(z)$ l'hypothèse $H$ est rejetée si

$$d^+_{n_1 n_2} > d_{\alpha'}, \tag{2.35}$$

où le seuil $d_{\alpha'}$ est donné pour un niveau significatif $\alpha'$ donné à partir de la condition (pour $N \gg 1$)

$$d_{\alpha'} = \left(\frac{1}{2N} \ln \frac{1}{\alpha'}\right)^{\frac{1}{2}}. \tag{2.35'}$$

Un autre critère à deux échantillons, proposé par Wilcoxon et étudié par Mann et Witni, est basé sur le comptage du *nombre d'inversions*. A cet effet on forme des deux échantillons x et y une série variationnelle $y^{(1)}, y^{(2)}, x^{(1)}, y^{(3)}, x^{(2)}, \ldots$ Si dans cette suite un $x^{(i)}$ donné est précédé par $s$ éléments $y^{(1)}, \ldots, y^{(s)}$ de l'échantillon y, on a $s$ inversions. Le nombre total d'inversions $U$ est égal à la somme des inversions formées par tous les éléments du premier échantillon avec les éléments du second. Il est évident

que

$$U = \sum_{j=1}^{n_2} \sum_{i=1}^{n_1} u(x_i - y_j). \tag{2.36}$$

On peut montrer (voir [4], page 337) que pour $n_1 + n_2 \geqslant 20$, $n_2 > 3$ on peut considérer que la distribution du nombre total d'inversions est normale avec une bonne approximation de paramètres

$$m_1\{U\} = \frac{n_1 n_2}{2}, \quad M_2\{U\} = \frac{n_1 n_2}{12}(n_1 + n_2 + 1). \tag{2.36'}$$

S'il est possible d'utiliser une distribution asymptotiquement normale de la statistique $U$, la règle de vérification de l'hypothèse $H$ d'appartenance de deux échantillons à une même distribution d'après le critère de Wilcoxon se formule comme suit.

Pour l'alternative $F_{1x}(z) \neq F_{1y}(z)$, l'hypothèse $H$ est rejetée si, pour un niveau significatif donné $\alpha$, on a

$$\left| \frac{U - m_1\{U\}}{\sqrt{M^2\{U\}}} \right| > x_{\alpha/2}, \tag{2.37}$$

où $x_{\alpha/2}$ est l'$\alpha/2$-quantile de la loi normale.

Pour une alternative unilatérale $F_{1x}(z) > F_{1y}(z)$ l'hypothèse $H$ est rejetée, si pour un niveau significatif donné $\alpha$ on a

$$\frac{U - m_1\{U\}}{\sqrt{M_2\{U\}}} > x_\alpha. \tag{2.37'}$$

Les grandeurs $m_1\{U\}$ et $M_2\{U\}$ dans (2.37) et (2.37″) sont données par les formules (2.36′).

## 2.2. STATISTIQUES UTILISÉES DANS LES ALGORITHMES NON PARAMÉTRIQUES DE DÉTECTION DES SIGNAUX

**2.2.1. Statistiques de signe.** A partir d'un échantillon d'observations indépendant on peut réaliser la synthèse des algorithmes non paramétriques de détection qui laisse inchangée la probabilité de fausse alarme pour une distribution quelconque du bruit stationnaire et une *taille arbitraire* de l'échantillon. Avec ces algorithmes, la probabilité de perte du signal sera évidemment supérieure à la valeur théorique et dépendra de la distribution du bruit. Nous allons envisager des algorithmes non paramétriques de détection des signaux utilisant des tests de signe, de rang et mixtes.

Soit $\mathbf{x} = (x_1, \ldots, x_n)$ l'échantillon observé. Introduisons la *fonction de signe*

$$\operatorname{sgn} x = \frac{x}{|x|} = \begin{cases} 1, & x \geqslant 0, \\ -1, & x < 0. \end{cases} \tag{2.38}$$

Nous appellerons *vecteur de signe* d'un échantillon le vecteur sgn x = (sgn $x_1$, . . ., sgn $x_n$). Nous appellerons *statistique de signe* une fonction arbitraire du vecteur de signe, et *algorithme de signe*, l'algorithme utilisant les seuls signes des éléments de l'échantillon.

Si la distribution du bruit stationnaire est symétrique par rapport au zéro, les nombres de signes positifs et négatifs dans l'échantillon sont équiprobables indépendamment de la forme de bruit. En présence d'un signal constant (positif) noyé dans un bruit, la probabilité des signes positifs dans l'échantillon devient supérieure à la probabilité des signes négatifs, ce qui permet de détecter le signal. Ainsi, les algorithmes de signe sont sensibles à la composante constante du signal.

Parfois au lieu de la fonction de signe (2.38) on utilise la fonction du saut unitaire (2.1'). Le vecteur u (x) de composantes [$u(x_1)$, . . . . . ., $u(x_n)$] est dit vecteur de signe *positif* de l'échantillon. La fonction de signe est liée d'une manière univoque à la fonction du saut unitaire, en effet

$$2\,u(x) = \operatorname{sgn} x + 1. \tag{2.39}$$

**2.2.2. Statistiques d'ordre.** Rangeons les éléments de l'échantillon x = ($x_1$, . . ., $x_n$) par ordre croissant, i.e. $x_k \leqslant x_j$ si $k < j$. On obtient alors un *échantillon ordonné* $x^{(1)}$, . . ., $x^{(n)}$ (série variationnelle). Le vecteur $x^{(\,)}$ dont les éléments coïncident avec les éléments de la série variationnelle est dit *vecteur des statistiques d'ordre*, et les éléments de ce vecteur, *statistiques d'ordre*. Pour des échantillons issus d'une distribution de densité continue, la probabilité de confusion de deux ou plus valeurs échantillonnées et par conséquent des statistiques d'ordre est nulle.

On peut facilement trouver la fonction de répartition $F_1^{(i)}(x)$ de la statistique d'ordre $x^{(i)}$ pour un échantillon homogène indépendant de taille $n$ issu de la distribution $F_1(x)$. Comme la probabilité de trouver dans l'échantillon indépendant ($x_1$, . . ., $x_n$) exactement $k$ éléments ne dépassant pas un seuil donné $x$ est égale à

$$\binom{n}{k} [F_1(x)]^k [1 - F_1(x)]^{n-k},$$

on a

$$F_1^{(i)}(x) = P\{x^{(i)} \leqslant x\} = \sum_{k=i}^{n} P\{x^{(k)} \leqslant x \leqslant x^{(k+1)}\} =$$

$$= \sum_{k=i}^{n} \binom{n}{k} [F_1(x)]^k [1 - F_1(x)]^{n-k}, \tag{2.40}$$

$$i = 1, \ldots, n; \quad x^{(n+1)} = \infty.$$

En dérivant par rapport à $x$ le membre de droite de (2.40) on obtient l'expression de la densité de probabilité de la statistique d'ordre $x^{(i)}$

$$w_1^{(i)}(x)=\frac{d}{dx}F_1^{(i)}(x)=w_1(x)\sum_{k=i}^{n}\binom{n}{k}\{k[F_1(x)]^{k-1}[1-F_1(x)]^{n-k}-$$

$$-(n-k)[F_1(x)]^k[1-F_1(x)]^{n-k-1}\}=$$

$$=nw_1(x)\left\{\sum_{k=i}^{n}\binom{n-1}{k-1}[F_1(x)]^{k-1}[1-F_1(x)]^{n-k}-\right.$$

$$\left.-\sum_{k=i}^{n}\binom{n-1}{k}[F_1(x)]^k[1-F_1(x)]^{n-k-1}\right\}=$$

$$=n\binom{n-1}{i-1}[F_1(x)]^{i-1}[1-F_1(x)]^{n-i}w_1(x). \qquad (2.40')$$

La densité de probabilité simultanée de $r$ statistiques d'ordre $y_i=x^{(s_i)}$, $1\leqslant s_1\leqslant\ldots\leqslant s_r\leqslant n$ est égale à (voir [5])

$$w_r(y_1,\ldots,y_r)=n![(s_1-1)!(s_2-s_1-1)!\ldots(n-s_r)!]^{-1}\times$$

$$\times F_1^{s_1-1}(y_1)[F_1(y_2)-F_1(y_1)]^{s_2-s_1-1}\ldots[1-F_1(y_r)]^{n-s_r}\prod_{i=1}^{r}w_1(y_i). \qquad (2.40'')$$

En vertu de (2.40″) et (2.40′) l'ensemble des statistiques d'ordre $x^{(1)},\ldots,x^{(n)}$ est une *suite markovienne* simple, en effet

$$w(y_{r+1}\mid y_r,\ldots,y_1)=\frac{w_{r+1}(y_1,\ldots,y_{r+1})}{w_r(y_1,\ldots,y_r)}=$$

$$=(n-r)w_1(y_{r+1})\frac{1-F_1^{n-r-1}(y_{r+1})}{1-F_1^{n-r}(y_r)}=\frac{w_2(y_r,y_{r+1})}{w_1(y_r)}=w(y_{r+1}\mid y_r).$$

**2.2.3. Statistiques de rang.** On appelle *rang* $R_i$ de l'élément $x_i$ d'un échantillon le numéro d'ordre de cet élément dans la série variationnelle ou, en d'autres termes, le nombre d'éléments de l'échantillon x inférieurs ou égaux à $x_i$. Par conséquent, à la valeur échantillonnée $x_i$ correspond la statistique $x^{(R_i)}$ de la série variationnelle.

On appelle *vecteur de rang* $\mathbf{R}(\mathbf{x})=(R_1,\ldots,R_n)$ de l'échantillon x la permutation des nombres $1,2,\ldots,n$ obtenue en remplaçant les éléments de l'échantillon par leur rang. On appelle *statistique de rang* une fonction arbitraire du vecteur de rang. *L'algorithme de rang* est réalisé comme un test de comparaison d'une certaine statistique de rang à un seuil.

On peut reconstituer l'échantillon initial $\mathbf{x}$ si l'on connaît le vecteur $\mathbf{x}^{(\ )}$ des statistiques d'ordre et le vecteur de rang $\mathbf{R}$. Pris séparément, chacun de ces deux vecteurs est une transformation non linéaire irréversible de l'échantillon initial. Pour un échantillon homogène indépendant $\mathbf{x}$ les vecteurs aléatoires $\mathbf{x}^{(\ )}$ et $\mathbf{R}$ sont indépendants.

On peut représenter le rang $R_i$ de l'élément $x_i$ d'un échantillon de taille $n$ à l'aide de la fonction du saut unitaire $u(x)$ ou de la fonction de signe comme suit:

$$R_i = \sum_{k=1}^{n} u(x_i - x_k), \quad i = 1, \ldots, n, \tag{2.41}$$

$$R_i = \frac{1}{2} \sum_{k=1}^{n} \operatorname{sgn}(x_i - x_k) + \frac{n}{2}, \quad i = 1, \ldots, n. \tag{2.42}$$

En vertu de (2.41) et (2.42), les rangs sont des statistiques de signe des différences des valeurs échantillonnées.

Pour un échantillon homogène indépendant la fonction de vraisemblance est *invariante* par rapport au groupe des permutations des arguments. Par conséquent, pour un tel échantillon tous les vecteurs de rang sont *équiprobables*, quelle que soit la distribution à laquelle appartient l'échantillon. Le nombre total de vecteurs de rang possibles correspondant à un échantillon de taille $n$ est égal au nombre de permutations de $n$ nombres, soit $n!$. Autrement dit, l'espace des vecteurs de rang des échantillons est formé par $n!$ points discrets de l'espace euclidien à $n$ dimensions. La probabilité pour le vecteur de rang $\mathbf{R}$ de l'échantillon observé de se trouver dans un point quelconque $\mathbf{r}_i$, $i = 1, \ldots, n!$, de cet ensemble discret est égale à $1/n!$, c'est-à-dire pour chaque $i = 1, \ldots, n!$ et pour une distribution quelconque homogène indépendante de l'échantillon on a

$$P\{\mathbf{R} = \mathbf{r}_i | H\} = 1/n!. \tag{2.43}$$

En vertu de (2.43) un algorithme de rang utilisé pour la détection d'un signal préserve la probabilité de fausse alarme pour un bruit stationnaire indépendant de distribution quelconque. Ainsi, un algorithme de rang est non paramétrique par rapport à un bruit stationnaire indépendant.

En présence d'un signal variable dans le temps, les rangs cessent d'être équiprobables car l'échantillon n'est plus homogène *). Le rangement des éléments de l'échantillon par ordre de grandeur est déterminé alors par la forme du signal, ce qui permet de le détec-

*) Par exemple, pour $n = 2$ on a pour un échantillon indépendant homogène

$$w_2(x_1, x_2) = w_{11}(x_1)\, w_{12}(x_2) \neq w_{11}(x_2)\, w_{12}(x_1).$$

ter. Ainsi, l'algorithme de rang est sensible à la non-stationnarité introduite par un signal variable dans le bruit stationnaire.

Pour trouver la fonction de répartition du vecteur de rang d'un échantillon non homogène appartenant à la distribution $w_n(\mathbf{x}|K)$ (par exemple, du mélange d'un bruit stationnaire et d'un signal variable dans le temps), il y a lieu de calculer l'intégrale

$$P\{\mathbf{R}=\mathbf{r}_i \mid K\}=\int\limits_{A_i}\cdots\int w_n(x_1,\ldots,x_n \mid K)\,dx_1\ldots dx_n,$$

où $A_i$ est le domaine contenant les points de l'espace échantillonné, auxquels correspond lors de la mise en ordre le vecteur donné $\mathbf{r}_i = (r_1^i, \ldots, r_n^{(i)})$. Cette intégrale, comme l'a montré Hoeffding, peut être ramenée à la forme [5]:

$$P\{\mathbf{R}=\mathbf{r}_i \mid K\}=\frac{1}{n!}\,m_1\left\{\frac{w_n[x^{(r_1^{(i)})},\ldots,x^{(r_n^{(i)})}\mid K]}{w_n[x^{(r_1^{(i)})},\ldots,x^{(r_n^{(i)})}\mid H]}\,\middle|\,H\right\}, \qquad (2.43')$$

où $w(\cdot \mid H)$ est la densité d'un échantillon homogène et $w(\cdot \mid H) > 0$ pour $w(\cdot \mid K) > 0$. Malheureusement l'utilisation pratique de la formule (2.43′), sauf quelques cas spéciaux, se heurte à des difficultés de calcul insurmontables.

Remarquons également que le vecteur de rang d'un bruit stationnaire est *invariant* par une transformation non inertielle de l'échantillon, i.e.

$$\mathbf{R}(\mathbf{x}) = \mathbf{R}(\mathbf{f}_\mathbf{x}), \qquad (2.43'')$$

où $\mathbf{f}_\mathbf{x}$ est le vecteur de composantes $f(x_1), \ldots, f(x_n)$, et $f(x)$ une fonction non décroissante, car une telle transformation ne change pas l'ordre relatif des éléments de l'échantillon $\mathbf{x}$. En vertu de (2.43″), après la transformation non linéaire mentionnée l'algorithme de rang reste *non paramétrique* vis-à-vis d'un bruit stationnaire indépendant.

Dans certains problèmes on utilise le vecteur des valeurs absolues des observations $|\mathbf{x}| = (|x_1|, \ldots, |x_n|)$ et le *vecteur des rangs positifs* $\mathbf{R}^+ = (R_1^+, \ldots, R_n^+)$ donnant les numéros d'ordre des éléments de la série variationnelle $|x|^{(1)}, \ldots, |x|^{(n)}$ de l'échantillon des valeurs absolues des observations. Il est évident que $x_i = |x|^{(R_i)} \operatorname{sgn} x_i$. Tout comme (2.41), on peut écrire $R_i^+$ sous la forme d'une somme, soit

$$R_i^+=\sum_{k=1}^{n} u(|x_i|-|x_k|), \quad i=1,\ldots,n.$$

L'*algorithme mixte* utilise la *statistique de signe et de rang* dépendant tant du vecteur des rangs $\mathbf{R}$ que du vecteur des signes $\operatorname{sgn}\mathbf{x}$ de l'échantillon observé.

**2.2.4. Volume des calculs que nécessite la réalisation des algorithmes de rang.** Les dispositifs réalisant les algorithmes de rang et mixtes effectuent un traitement discret des observations. Pour prendre une décision on utilise non pas les grandeurs observées, ni même les valeurs quantifiées en amplitude des processus observés, mais seulement la suite des nombres naturels, de l'unité à $n$, liées à l'observation.

Formons pour chaque élément $x_i$ de l'échantillon $x_1, \ldots, x_n$ les $n$ différences $x_i - x_j$, $j = 1, \ldots, n$. En supposant (pour des distributions continues) $x_i - x_j \neq 0$, $i \neq j$, on voit que le rang de l'élément $x_i$ est égal au nombre des différences $x_i - x_j > 0$ (voir (2.41)). Composons la matrice dont les éléments sont $u(x_i - x_j)$. C'est une matrice $n \times n$ aux éléments diagonaux nuls. Le nombre d'éléments « 1 » dans la ligne $i$ est égal au rang de l'élément $x_i$. Ainsi, pour calculer le vecteur de rang **R** d'un échantillon de taille $n$ il y a lieu de former $n^2$ différences et d'effectuer $n$ sommations des lignes de la matrice.

## 2.3. ALGORITHMES DE SIGNE DE DÉTECTION DES SIGNAUX NOYÉS DANS DES BRUITS INDÉPENDANTS

**2.3.1. Détecteur de signe unilatéral du signal constant.** Nous allons envisager le problème de la détection d'un signal constant $s(t) \equiv a$ noyé dans un bruit stationnaire indépendant additif de moyenne nulle et de distribution *symétrique* $w_1(x)$ *). Ce problème consiste à vérifier l'hypothèse $H$ sur la symétrie par rapport à l'origine de la fonction de répartition du bruit (i.e. vérifier l'hypothèse selon laquelle cette fonction est paire) contre l'alternative $K$ que la fonction de répartition n'est pas symétrique par rapport à l'origine. L'hypothèse $H$ entraîne l'affirmation que la médiane de la distribution est nulle, i.e.

$$p = P\{x > 0 \mid H\} = 1/2, \tag{2.44}$$

et l'alternative $K$ entraîne que la médiane n'est pas nulle, i.e.

$$p = P\{x > 0 \mid K\} \neq 1/2. \tag{2.45}$$

La réciproque n'est en général pas vraie, c'est-à-dire que (2.44) *n'entraîne obligatoirement pas la symétrie.* Pour les distributions symétriques la médiane et la moyenne coïncident. Supposons tout d'abord que $a > 0$, c'est-à-dire que $p > 1/2$. Le nombre de valeurs

*) On suppose également que la fonction de répartition $F_1(x) = \int_{-\infty}^{x} w_1(u)\,du$ est continue.

positives et négatives des échantillons pour $n$ grand sera voisin de $n/2$ en l'absence de signal et supérieur à $n/2$ en présence d'un signal positif.

Pour la détection d'un signal constant positif (pour $p > 1/2$) noyé dans un bruit stationnaire additif à distribution symétrique il est tout naturel d'utiliser l'algorithme de signe : le signal est présent (l'hypothèse $H$ est rejetée) si pour un échantillon indépendant d'observations $x_1, \ldots, x_n$ on a

$$\sum_{i=1}^{n} \operatorname{sgn} x_i \geqslant C_1. \tag{2.46}$$

Compte tenu de la relation existant entre les fonctions sgn $x$ et $u(x)$ (voir (2.39)), on peut formuler la règle de choix de décision comme suit : le signal est présent (l'hypothèse $H$ est rejetée) si

$$\sum_{i=1}^{n} u(x_i) \geqslant C, \tag{2.47}$$

et il n'y a pas de signal si l'inégalité inverse de (2.47) est vérifiée. Le seuil $C$ est déterminé par la probabilité donnée de fausse alarme. L'algorithme de signe (2.47) est réalisé par un limiteur unilatéral parfait, un sommateur et un dispositif de comparaison à seuil (fig. 2.1).

La somme dans le membre de gauche de (2.47), égale au nombre des valeurs positives dans un échantillon indépendant de taille $n$,

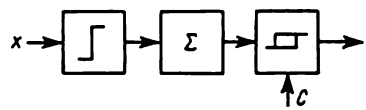



Fig. 2.1. Schéma fonctionnel du détecteur de signe d'un signal constant.

suit une loi de répartition des probabilités binomiale de paramètres $(n, 1/2)$ si l'hypothèse $H$ est vérifiée et de paramètres $(n, p)$ si c'est l'alternative $K$ qui est vérifiée. En utilisant l'expression de la fonction intégrale de la distibution binomiale on trouve pour une probabilité donnée de fausse alarme $\alpha$ la grandeur du seuil $C$ et la probabilité de perte du signal $\beta$. Il découle de (2.47) qu'il existe toujours un $\alpha^* \leqslant \alpha$ tel que

$$\alpha^* = P\left\{\sum_{i=1}^{n} u(x_i) \geqslant [C]+1 \mid H\right\} = \sum_{k=[C]+1}^{n} \binom{n}{k} (1/2)^n =$$
$$= 1 - \mathrm{I}_{1/2}(n-[C], [C]+1), \tag{2.48}$$

où $I_q(a, b)$ est le rapport de la fonction bêta incomplète à la fonction bêta complète, $[C]$ la partie entière de $C$. L'équation (2.48) donne un *seuil constant* pour des distributions quelconques du bruit pour une valeur donnée de la probabilité de fausse alarme *).

La probabilité de perte de signal est

$$\beta = P\left\{\sum_{i=1}^{n} u(x_i) \leqslant [C] \mid K\right\} = \sum_{k=0}^{[C]} \binom{n}{k} p^k (1-p)^{n-k} =$$

$$= I_{1-p}(n-[C], [C]+1). \qquad (2.49)$$

En vertu de (2.48) et (2.49), pour $a > 0$, i.e. pour $p > 1/2$, la règle (2.47) est *non biaisée* pour une probabilité de fausse alarme égale à $\alpha$ *. En effet, pour $q = 1 - p < 1/2$ l'inégalité (voir tome I, (1.22))

$$B_{1/2}(n-m, m+1) = \int_0^{1/2} z^{n-m-1}(1-z)^m\, dz > \int_0^{q} z^{n-m-1}(1-z)^m\, dz =$$

$$= B_q(n-m, m+1)$$

entraîne

$$I_{1/2}(n-[C], [C]+1) > I_q(n-[C], [C]+1)$$

et par conséquent

$$1 - \beta > \alpha\ *.$$

Pour un échantillon de taille importante la distribution binomiale peut être approximée par une loi normale (voir tome I, 1.2.2) de moyenne $np$ et de variance $np$ (1 — p). Les formules de la probabilité d'erreurs de première et de seconde espèce pour $n \gg 1$ peuvent alors s'écrire sous la forme

$$\alpha \sim 1 - F\left(\frac{C - n/2}{\sqrt{n/4}}\right), \qquad (2.50)$$

$$\beta \sim F\left(\frac{C - np}{\sqrt{np(1-p)}}\right), \qquad (2.51)$$

où $F(x)$ est l'intégrale de Laplace.

Pour une probabilité donnée de l'erreur de première espèce $\alpha$ le seuil $C$ est donné par (2.50), soit

$$C = (x_\alpha \sqrt{n} + n)/2, \qquad (2.51')$$

où $x_\alpha$ est le $\alpha$-quantile de la loi normale. En substituant (2.51')

*) Plus exactement, l'équation (2.48) détermine $[C]$. Il se peut qu'il n'existe pas de nombre entier $[C]$ pour lequel le membre de droite de (2.48) soit exactement égal à $\alpha$. Dans A.3.1 on trouvera une table des grandeurs $[C]$ empruntée dans [4].

dans (2.50) on obtient pour $n \gg 1$

$$x_{1-\beta} = \frac{1}{2\sqrt{p(1-p)}}\left(x_\alpha - (2p-1)\sqrt{n}\right). \tag{2.52}$$

(2.52) permet de voir que pour une règle non biaisée ($a > 0$) pour $n \to \infty$ la probabilité d'erreurs de seconde espèce $\beta \to 0$.

Si $a < 0$ (i.e. $p < 1/2$), la règle non biaisée sera celle qui implique le rejet de l'hypothèse $H$ lorsque l'inégalité inverse de (2.47) est vérifiée.

On a alors

$$\alpha^* = P\left\{\sum_{i=1}^{n} u(x_i) < [C] \mid H\right\} = \mathrm{I}_{1/2}(n-[C], [C]+1), \tag{2.53}$$

$$\beta = P\left\{\sum_{i=1}^{n} u(x_i) \geqslant [C]+1 \mid K\right\} = 1 - \mathrm{I}_{1-p}(n-[C], [C]+1) \tag{2.54}$$

et pour $p < 1/2$ en vertu de (2.53) et (2.54) on a $1 - \beta > \alpha^*$.

**2.3.2. Détecteur de signe bilatéral d'un signal constant.** Pour un signal de signe quelconque, i.e. pour $p \neq 1/2$ quelconque, on peut formuler la règle *bilatérale* non biaisée de choix d'une décision en vertu de laquelle l'hypothèse $H$ sur la symétrie de la distribution est rejetée si le nombre total d'éléments positifs de l'échantillon ou le nombre total d'éléments négatifs dépasse un certain seuil, c'est-à-dire

$$\sum_{i=1}^{n} u(x_i) \geqslant C \quad \text{ou} \quad n - \sum_{i=1}^{n} u(x_i) > C. \tag{2.55}$$

La probabilité de fausse alarme est alors égale à (comparer avec (2.48))

$$\begin{aligned}\alpha^* &= P\left\{\sum_{i=1}^{n} u(x_i) \geqslant [C]+1 \mid H\right\} + P\left\{\sum_{i=1}^{n} u(x_i) < n-[C] \mid H\right\} = \\ &= 1 - \mathrm{I}_{1/2}(n-[C], [C]+1) + \mathrm{I}_{1/2}([C]+1, n-[C]) = \\ &= 2\left[1 - \mathrm{I}_{1/2}(n-[C], [C]+1)\right].\end{aligned} \tag{2.56}$$

La probabilité de détection correcte est

$$\begin{aligned}1-\beta &= P\left\{\sum_{i=1}^{n} u(x_i) > [C] \mid K\right\} + P\left\{\sum_{i=1}^{n} u(x_i) < n-[C] \mid K\right\} = \\ &= \sum_{k=[C]+1}^{n} \binom{n}{k} p^k (1-p)^{n-k} + \sum_{k=0}^{n-[C]-1} \binom{n}{k} p^k (1-p)^{n-k} = \\ &= 1 - \mathrm{I}_{1-p}(n-[C], [C]+1) + \mathrm{I}_{1-p}([C]+1, n-[C]).\end{aligned} \tag{2.57}$$

Il est facile de montrer que la règle (2.55) est non biaisée, en effet

$$1-\beta=1-I_{1-p}(n-[C],\ [C]+1)+I_{1-p}([C]+1,\ n-[C])>$$
$$>1-I_{1/2}(n-[C],\ [C]+1)+I_{1/2}([C]+1,\ n-[C])=\alpha^*.$$

Pour $n \gg 1$, d'une manière analogue à (2.50) et (2.51) on trouve

$$\alpha \sim 2\left[1-F\left(\frac{2C-n}{\sqrt{n}}\right)\right], \tag{2.58}$$

$$\beta \sim F\left(\frac{C-np}{\sqrt{np(1-p)}}\right)-F\left(\frac{n(1-p)-C}{\sqrt{np(1-p)}}\right). \tag{2.59}$$

Pour une probabilité donnée de fausse alarme $\alpha$ le seuil $C$ est trouvé à partir de (2.58)

$$C=(x_{\alpha/2}\sqrt{n}+n)/2. \tag{2.60}$$

En portant (2.60) dans (2.59) on trouve

$$\beta \sim F\left[\frac{x_{\alpha/2}-(2p-1)\sqrt{n}}{2\sqrt{p(1-p)}}\right]-F\left[\frac{-x_{\alpha/2}-(2p-1)\sqrt{n}}{2\sqrt{p(1-p)}}\right]. \tag{2.61}$$

**2.3.3. Efficacité relative du détecteur de signe unilatéral.** Comparons l'algorithme de signe unilatéral (2.47) avec l'algorithme optimal pour la détection d'un signal positif constant noyé dans un bruit normal additif de moyenne nulle et de variance donnée.

La règle optimale (uniformément la plus puissante) suivant le critère de Neyman-Pearson de détection d'un signal positif constant noyé dans un bruit normal additif d'après un échantillon indépendant d'observations $(x_1, \ldots, x_n)$ consiste à comparer les sommes des valeurs échantillonnées à un seuil déterminé pour une valeur donnée de la probabilité de fausse alarme (voir tome II, 1.3.1). Supposons que l'algorithme linéaire unilatéral

$$\sum_{k=1}^{n} x_k \geqslant C, \tag{2.62}$$

optimal pour un bruit normal, soit utilisé pour la détection d'un signal positif $a>0$ noyé dans un bruit additif de moyenne nulle, ayant une distribution $w_1(x)$ symétrique quelconque de variance $\sigma^2<\infty$. En vertu du théorème de la limite centrale pour un échantillon indépendant issu d'une distribution quelconque de variance finie, la statistique du membre de gauche de (2.62) est asymptotiquement normale (pour $n \gg 1$) de paramètres $(0, n\sigma^2)$ en l'absence de signal et $(na, n\sigma^2)$ en présence du signal. Les probabilités de fausse

alarme et de perte de signal pour $n \gg 1$ sont respectivement

$$\alpha = P\left\{\sum_{k=1}^{n} x_k \geqslant C \mid a=0\right\} \sim 1 - F\left(\frac{C}{\sigma\sqrt{n}}\right), \tag{2.63}$$

$$\beta = P\left\{\sum_{k=1}^{n} x_k < C \mid a>0\right\} \sim F\left(\frac{C-na}{\sigma\sqrt{n}}\right). \tag{2.64}$$

Pour une probabilité donnée de fausse alarme $\alpha$ on obtient à partir de (2.63) le seuil

$$C = x_\alpha \sigma \sqrt{n} \tag{2.65}$$

et en substituant (2.65) dans (2.64) on obtient (voir tome II, (1.77′))

$$x_{1-\beta} = x_\alpha - \left(\frac{a}{\sigma}\right)\sqrt{n}. \tag{2.66}$$

Fixons maintenant les valeurs $\alpha$, $\beta$ des probabilités d'erreurs, et soient $n_1$ et $n_2$ les tailles des échantillons nécessaires pour obtenir les valeurs données des erreurs dans le cas de détection du signal d'après l'algorithme de signe (2.47) et l'algorithme linéaire (2.62) respectivement. En vertu de (2.52) et (2.66) on obtient immédiatement

$$x_{1-\beta} = \frac{1}{2\sqrt{p(1-p)}}\left(x_\alpha - (2p-1)\sqrt{n_1}\right), \tag{2.67}$$

$$x_{1-\beta} = x_\alpha - \frac{a}{\sigma}\sqrt{n_2}. \tag{2.67′}$$

Il est facile de voir que (2.67) et (2.67′) correspondent à la règle générale (1.71), car dans le cas envisagé on a

$$a_{nH}^{(1)} = \frac{n}{2}, \quad a_{nK}^{(1)} = np,$$
$$a_{nH}^{(2)} = 0, \quad a_{nK}^{(2)} = na,$$
$$\sigma_{nH}^{(1)} = \frac{\sqrt{n}}{2}, \quad \sigma_{nK}^{(1)} = \sqrt{np(1-p)},$$
$$\sigma_{nH}^{(2)} = \sigma_{nK}^{(2)} = \sqrt{n}\,\sigma.$$

Proposons-nous de trouver le coefficient d'efficacité asymptotique relative $\rho$ de l'algorithme de signe par rapport à l'algorithme linéaire, en utilisant à cet effet la formule générale (1.72) et en tenant compte du fait que les distributions des algorithmes mentionnés sont asymptotiquement normales.

Puisque nous avons supposé que la fonction de distribution $w_1(x)$ du bruit additif est symétrique, nous avons

$$p = P\{x>0 \mid a>0\} = \int_0^\infty w_1(x-a)\,dx = \int_{-a}^\infty w_1(y)\,dy =$$
$$= \frac{1}{2} + \int_0^a w_1(x+a)\,dx = \frac{1}{2} + a w_1(0) + o(a), \tag{2.68}$$

$$4p\,(1-p) = 1 + o\,(a), \tag{2.68'}$$

et à partir de (1.70′) il vient

$$\varepsilon_1 = \lim_{n\to\infty} n w_1\,(0)/(n/2) = 2w_1'(0), \tag{2.69}$$

$$\varepsilon_2 = \lim_{n\to\infty} \frac{n}{n\sigma} = \frac{1}{\sigma}. \tag{2.69'}$$

En substituant (2.69) et (2.69′) dans (1.72) on obtient l'expression du coefficient d'efficacité asymptotique relative

$$\rho = 4\sigma^2 w_1^2\,(0). \tag{2.70}$$

La formule (2.70) peut évidemment être déduite directement à partir de (2.67), (2.67′) et (2.68) en posant $a\sqrt{n_1} = \gamma_1$, $a\sqrt{n_2} = \gamma_2$. En négligeant les infiniment petits $o\,(a)$, les racines de ces équations sont

$$\gamma_1 = \frac{x_\alpha - x_{1-\beta}}{2w_1\,(0)}, \qquad \gamma_2 = \sigma\,(x_\alpha - x_{1-\beta}),$$

ce qui permet de calculer le rapport $(\gamma_2/\gamma_1)^2$.

Dans l'hypothèse que la distribution du bruit est normale, on a $w_1\,(0) = (2\pi\sigma^2)^{-1/2}$, alors en vertu de (2.69) on trouve

$$\rho = 2/\pi \sim 0{,}64, \tag{2.71}$$

donc le détecteur de signe (pour des tailles importantes de l'échantillon et un signal faible) est presque deux fois moins efficace que le détecteur linéaire optimal pour un bruit normal.

Mais il en va tout autrement pour une distribution du bruit autre que la loi normale. Ainsi pour un *bruit laplacien*

$$w_1\,(x) = \frac{\lambda}{2} e^{-\lambda|x|}, \qquad \lambda > 0, \tag{2.72}$$

on a

$$w_1\,(0) = \lambda/2, \quad \sigma^2 = 2/\lambda^2.$$

En vertu de (2.70), on obtient alors

$$\rho = 2, \tag{2.72'}$$

i.e. le détecteur de signe est (asymptotiquement) deux fois plus efficace que le détecteur linéaire. Mais pour un bruit uniforme pour lequel

$$w_1\,(x) = 1/2a, \quad |\,x\,| \leqslant a, \quad \sigma^2 = a^2/3, \tag{2.73}$$

on obtient à partir de (2.70)

$$\rho = 1/3, \tag{2.73'}$$

et pour un bruit sinusoïdal de phase aléatoire, de distribution

$$w_1(x) = \frac{1}{\pi\sqrt{a^2 - x^2}}, \quad |x| \leqslant a, \quad \sigma^2 = \frac{a^2}{2}, \tag{2.74}$$

on a

$$\rho = 2/\pi^2 \sim 0{,}2, \tag{2.74'}$$

c'est-à-dire que dans ces cas le détecteur linéaire est bien plus efficace que le détecteur de signe.

En vertu de (2.70), pour un bruit de distribution symétrique, pour lequel $w_1(0) = 0$, le coefficient d'efficacité asymptotique relative de l'algorithme de signe par rapport à l'algorithme linéaire est égal à zéro. A titre d'exemple on peut citer la distribution bimodale suivante

$$w_1(x) = (1/2)\,|x|\,e^{-|x|}.$$

Remarquons (voir tome II, 1.5.1) qu'en vertu de la règle optimale suivant le critère de Neyman-Pearson de détection d'un signal positif noyé dans un bruit normal additif indépendant de moyenne nulle et de *variance inconnue* la décision que le signal est présent est prise dans le cas où pour une probabilité donnée de fausse alarme $\alpha$ on a

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} x_k \left[\frac{1}{n-1} \sum_{k=1}^{n} \left(x_k - \frac{1}{n} \sum_{i=1}^{n} x_i\right)^2\right]^{-1/2} \geqslant t_\alpha, \tag{2.75}$$

où $t_\alpha$ est l'$\alpha$-quantile de la loi de Student. Le carre du dénominateur dans le membre de gauche de (2.75) est une estimation non biaisée de la variance. Ainsi pour $n \to \infty$ la statistique (2.75) coïncide avec la statistique linéaire (2.62), avec $t_\alpha \to x_\alpha$, le seuil étant donné par (2.65). Ainsi, la formule (2.70) reste vraie également pour le coefficient d'efficacité asymptotique relative de l'algorithme de signe par rapport à l'algorithme (2.75).

**2.3.4. Efficacité relative du détecteur de signe bilatéral.** Nous allons maintenant comparer l'algorithme de signe bilatéral (2.55) avec l'algorithme optimal pour la détection d'un signal constant (de signe quelconque) noyé dans un bruit normal additif de moyenne nulle et de variance donnée.

On sait (voir tome II, 1.4.7) qu'en vertu de la règle optimale (non biaisée et uniformément la plus puissante) suivant le critère de Neyman-Pearson de détection d'un signal constant noyé dans un bruit normal additif d'après un échantillon indépendant d'observations $(x_1, \ldots, x_n)$, la décision sur la présence du signal est prise dans le cas où pour une probabilité donnée de fausse alarme $\alpha$ on a

$$\left|\sum_{i=1}^{n} x_i\right| \geqslant x_{\alpha/2}\sigma\sqrt{n}. \tag{2.76}$$

Supposons que l'algorithme linéaire bilatéral (2.76) est utilisé pour la détection d'un signal constant $a \neq 0$ noyé dans un bruit additif de moyenne nulle, de variance $\sigma^2$ et de distribution symétrique quelconque $w_1(x)$. La somme $\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} x_i$ étant pour l'alternative $K$ asymptotiquement normale de paramètres $(a\sqrt{n}/\sigma, 1)$, pour $n \gg 1$ la probabilité de perte de signal lors de la détection suivant l'algorithme (2.76) est

$$\beta = F\left(x_{\alpha/2} - \frac{a}{\sigma}\sqrt{n}\right) - F\left(-x_{\alpha/2} - \frac{a}{\sigma}\sqrt{n}\right). \qquad (2.77)$$

La probabilité de perte de signal lorsque l'on utilise l'algorithme (2.55), pour $n \gg 1$, est en vertu de (2.61) égale à

$$\beta = F\left[\frac{x_{\alpha/2} - (2p-1)\sqrt{n}}{2\sqrt{p(1-p)}}\right] - F\left[\frac{-x_{\alpha/2} - (2p-1)\sqrt{n}}{2\sqrt{p(1-p)}}\right]. \qquad (2.78)$$

En comparant (2.77), (2.78) avec les formules générales (1.74), (1.74′) on voit immédiatement que la grandeur $\rho$ dans le cas envisagé coïncide avec (2.70). Ainsi, le coefficient d'efficacité asymptotique relative de l'algorithme de signe bilatéral de détection d'un signal constant par rapport à l'algorithme linéaire bilatéral est égal au coefficient $\rho$ pour les algorithmes unilatéraux.

On sait (voir tome II, page 74) qu'en vertu de la règle optimale (non biaisée et uniformément la plus puissante) d'après le critère de Neyman-Pearson de détection d'un signal constant noyé dans un bruit normal additif de moyenne nulle et de *variance inconnue*, la décision sur la présence du signal est prise si pour une probabilité donnée de fausse alarme $\alpha$ on a

$$\frac{1}{\sqrt{n}}\left|\sum_{k=1}^{n} x_i\right|\left[\frac{1}{n-1}\sum_{k=1}^{n}\left(x_k - \frac{1}{n}\sum_{i=1}^{n} x_i\right)^2\right]^{-1/2} \geqslant t_{\alpha/2}, \qquad (2.79)$$

où $t_{\alpha/2}$ est l'$\alpha/2$-quantile de la loi de Student. Le carré du dénominateur dans le membre de gauche de (2.79) est une estimation non biaisée de la variance $\sigma^2$ de la distribution à laquelle appartient l'échantillon. C'est pourquoi pour $n \gg 1$, pour un échantillon homogène indépendant issu d'une distribution quelconque, (2.79) s'identifie à (2.76). Par conséquent, la formule (2.70) reste également vraie pour le coefficient d'efficacité asymptotique relative de l'algorithme de signe bilatéral par rapport à l'algorithme (2.79).

Rappelons que tous les algorithmes de signe que nous avons vus sont utilisés pour la détection d'un signal constant noyé dans un bruit additif à distribution *symétrique*. Il est évident qu'un détecteur de signe est inefficace dans le cas d'un signal constant

noyé dans un bruit à distribution unilatérale (par exemple, un bruit de Rayleigh). Qui plus est, si $w_1(0) = 0$ l'utilisation d'un algorithme de signe même pour un bruit à distribution symétrique est asymptotiquement non efficace. C'est le cas, par exemple, pour un bruit à distribution bimodale symétrique.

**2.3.5. Détecteur de signe d'un signal déterministe quelconque.** Le détecteur de signe peut être utilisé pour la détection d'un signal déterministe quelconque $\lambda s(t)$, $\lambda > 0$, noyé dans un bruit additif indépendant de moyenne nulle et de distribution symétrique $w_1(x)$.

Dans ce cas il faut vérifier l'hypothèse $H : \lambda = 0$ selon laquelle l'échantillon observé $x_1, \ldots, x_n$ appartient à la distribution du bruit contre l'alternative $K : \lambda > 0$ que l'échantillon appartient à la distribution du mélange du signal et du bruit. La règle de vérification de l'hypothèse utilisant un algorithme de signe se formule dans ce cas comme suit : l'hypothèse $H$ est rejetée (le signal est présent) si

$$\sum_{k=1}^{n} u(s_k x_k) \geqslant C, \tag{2.80}$$

où $s_k = s(t_k)$, $k = 1, \ldots, n$, et $\operatorname{sgn}[\lambda s(t)] = \operatorname{sgn} s(t)$ pour $\lambda > 0$. Le seuil $C$ est déterminé par une probabilité donnée $\alpha$ de

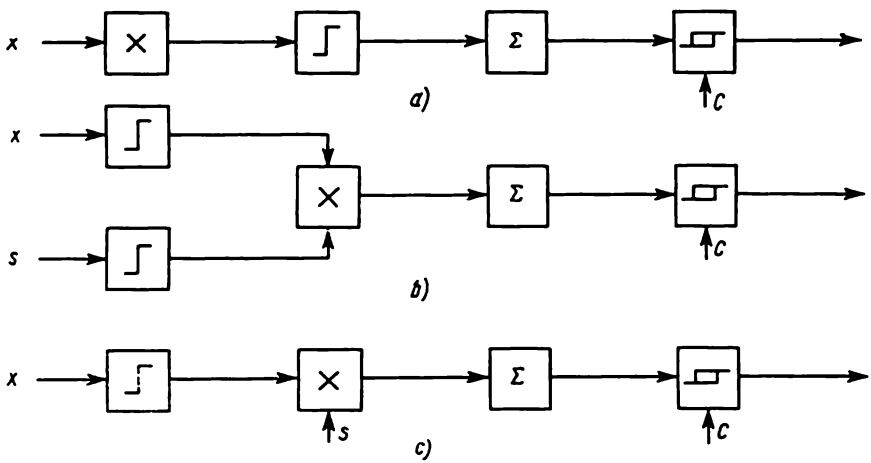



Fig. 2.2. Schémas fonctionnels des détecteurs de signe des signaux déterministes.

fausse alarme. Le schéma du détecteur de signe d'un signal déterministe est donné sur la figure 2.2, *a*.

L'inégalité (2.80) peut s'écrire comme suit (voir (2.39)):

$$\sum_{i=1}^{n} u(s_i x_i) = \frac{1}{2}\sum_{i=1}^{n} \operatorname{sgn}(s_i x_i) + \frac{n}{2} = \frac{1}{2}\sum_{i=1}^{n} \operatorname{sgn} s_i \operatorname{sgn} x_i + \frac{n}{2} \geqslant C. \quad (2.80')$$

A partir de (2.80') on voit que dans le schéma du détecteur de signe on peut effectuer la multiplication après la limitation du processus observé $x(t)$ et du signal déterministe $s(t)$ (fig. 2.2, *b*).

Dans une autre variante du détecteur de signe d'un signal déterministe on effectue d'abord une limitation du processus $x(t)$ et puis la multiplication par les valeurs du signal, obtenues à partir d'un oscillateur local (fig. 2.2, *c*). On prend la décision sur la présence du signal si

$$\sum_{k=1}^{n} s_k \operatorname{sgn} x_k \geqslant C \quad (2.80'')$$

ou (pour une variation correspondante du seuil)

$$\sum_{k=1}^{n} s_k u(x_k) \geqslant C. \quad (2.80''')$$

Pour un échantillon indépendant la somme des variables aléatoires dans (2.80) est asymptotiquement normale (voir 2.80') de paramètres

$$m_1\left\{\sum_{i=1}^{n} u(s_i x_i)\right\} = \frac{n}{2} + \frac{1}{2}\sum_{i=1}^{n} (2p_i - 1) \operatorname{sgn} s_i, \quad (2.81)$$

car

$$m_1\{\operatorname{sgn} x_i\} = 2p_i - 1, \quad p_i = P\{x_i \geqslant 0\} \quad (2.81')$$

et

$$M_2\left\{\sum_{i=1}^{n} u(s_i x_i)\right\} = \sum_{i=1}^{n} p_i(1 - p_i), \quad (2.82)$$

car

$$M_2\{\operatorname{sgn} x_i\} = 1 - (2p_i - 1)^2 = 4p_i(1 - p_i). \quad (2.82')$$

En l'absence de signal (hypothèse $H$) on a $p_i = 1/2$. Donc à partir de (2.81) et (2.82), on obtient pour $n \gg 1$ la probabilité de fausse alarme

$$\alpha \sim 1 - F\left(\frac{2C - n}{\sqrt{n}}\right), \quad (2.83)$$

d'où l'on trouve la valeur du seuil $C$ dans (2.80)

$$C = \frac{1}{2}(x_\alpha \sqrt{n} + n). \quad (2.83')$$

laquelle, comme on pouvait s'y attendre, coïncide avec (2.51').

Pour l'alternative $K$ (le signal est présent) on a (voir (2.68))

$$p_i = P\{x_i > 0\} = \frac{1}{2} + \lambda s_i w_1(0) + o(\lambda),$$

$$4p_i(1-p_i) = 1 - o(\lambda)$$

et par conséquent

$$m_1\left\{\sum_{i=1}^{n} u(s_i x_i) \mid K\right\} \sim \frac{n}{2} + \lambda w_1(0) \sum_{i=1}^{n} |s_i|, \quad |s_i| = s_i \operatorname{sgn} s_i, \tag{2.84}$$

$$M_2\left\{\sum_{i=1}^{n} u(s_i x_i) \mid K\right\} \sim \frac{n}{4}. \tag{2.84'}$$

La probabilité de perte de signal est

$$\beta \sim F\left[\frac{2C}{\sqrt{n}} - \sqrt{n}\,(1 + 2\lambda w_1(0)\, a_{|s|,\,n})\right], \tag{2.85}$$

où

$$a_{|s|,\,n} = \frac{1}{n} \sum_{i=1}^{n} |s_i|. \tag{2.85'}$$

A partir de (2.85) et (2.83') on trouve

$$x_{1-\beta} = x_\alpha - 2\lambda \sqrt{n}\, w_1(0)\, a_{|s|,\,n}. \tag{2.85''}$$

Remarquons que lorsqu'on utilise l'algorithme (2.80''), les quantiles $x_{1-\beta}$ et $x_\alpha$ sont liés par la relation

$$x_{1-\beta} = x_\alpha - 2\lambda \sqrt{n}\, w_1(0)\, \mathrm{W}_{sn}, \tag{2.85'''}$$

où

$$\mathrm{W}_{sn} = \frac{1}{n} \sum_{i=1}^{n} s_i^2. \tag{2.85''''}$$

**2.3.6. Efficacité relative du détecteur de signe d'un signal déterministe.** Nous allons comparer l'algorithme de signe (2.80) de détection d'un signal déterministe (de signe alterné) noyé dans un bruit additif avec l'algorithme linéaire

$$\sum_{k=1}^{n} s_k x_k \geqslant C, \tag{2.86}$$

qui, comme on sait (voir tome II, § 5.2), est un algorithme optimal, d'après le critère de Neyman-Pearson, de détection d'un signal

déterministe noyé dans un bruit normal additif d'après un échantillon indépendant d'observations.

Supposons que l'algorithme linéaire (2.86), optimal pour un bruit normal, est utilisé pour la détection d'un signal déterministe $\lambda s(t)$, $\lambda > 0$, noyé dans un bruit additif de moyenne nulle et de distribution symétrique quelconque de variance finie $\sigma^2$. En vertu du théorème de la limite centrale *), pour un échantillon indépendant appartenant à une distribution quelconque (pour $\sigma^2 < \infty$), la statistique dans le membre de gauche de (2.86) est asymptotiquement (pour $n \gg 1$) normale de paramètres $(0, n\sigma^2 W_{sn})$ quand il n'y a pas de signal, et de paramètres $(\lambda n\, W_{sn}, n\sigma^2 W_{sn})$ en présence du signal, car dans ce cas

$$m_1\{x_k\} = \lambda s_k, \quad s_k = s(t_k).$$

La probabilité de perte de signal $\beta$ et la probabilité donnée $\alpha$ de fausse alarme sont liées par (voir (2.66))

$$x_{1-\beta} = x_\alpha - \frac{\lambda}{\sigma}\sqrt{W_{sn}}\sqrt{n}. \tag{2.87}$$

Nous allons trouver le coefficient d'efficacité asymptotique relative de l'algorithme de signe (2.80) par rapport à l'algorithme linéaire. En comparant (2.87) avec (2.66), et (2.85) avec (2.52) nous voyons que le coefficient mentionné peut être obtenu à partir de (2.70) par multiplication par

$$V_s^2 = \lim_{n\to\infty} \frac{a^2_{|s|,n}}{W_{sn}} = \lim_{T\to\infty} \frac{\left(\int_0^T |s(t)|\,dt\right)^2}{\int_0^T s^2(t)\,dt}, \tag{2.88}$$

où $a_{|s|,n}$ et $W_{sn}$ sont donnés par (2.85′) et (2.85‴).

Ainsi, le coefficient cherché d'efficacité asymptotique relative est égal à

$$\rho = 4\sigma^2 w_1^2(0)\, V_s^2. \tag{2.89}$$

Remarquons que $V_s \leqslant 1$, avec $V_s = 1$ pour un signal constant.

Pour l'algorithme (2.80‴) en vertu de (2.85‴) et (2.87) on a

$$\rho = 4\sigma^2 w_1^2(0), \tag{2.89′}$$

ce qui coïncide avec (2.70). Le coefficient d'efficacité asymptotique relative de l'algorithme (2.80‴) par rapport à l'algorithme (2.80) est égal à $1/V_s^2 \geqslant 1$.

---

*) Avec la restriction supplémentaire suivante: pour $n \to \infty$ $\max_{1\leqslant k\leqslant n} s_k^2 / \sum_{k=1}^{n} s_k^2 \to 0$, ce qui est pratiquement toujours réalisé.

**2.3.7. Détecteur signe-quantiles d'un signal constant.** Supposons qu'en plus de savoir que la distribution du bruit additif est asymétrique par rapport à l'origine des coordonnées, on connaît $m$ quantiles de cette distribution $z_{q_j} = F_1^{-1}(q_j)$, $j = 1, \ldots, m$, où $F_1^{-1}(q)$ est la fonction inverse de la fonction de répartition du bruit *normée*, et $0 < q < 1$. Cette information a priori supplémentaire peut être utilisée pour augmenter l'efficacité du détecteur de signe (2.47) tout en conservant sa propriété d'être non paramétrique *).

Au lieu de calculer le nombre de signes positifs dans un échantillon indépendant on détermine le nombre des dépassements pour chaque quantile donné, puis la somme de ces nombres est comparée au seuil. Ainsi, on arrive à l'algorithme suivant pour la détection d'un signal constant (positif) $a$ noyé dans un bruit additif de variance $\sigma^2$ et de distribution symétrique: on prend la décision que le signal est présent si

$$\sum_{i=1}^{n} \sum_{j=1}^{m} u(x_i/\sigma - z_{q_j}) \geqslant C, \tag{2.90}$$

et la décision sur l'absence de signal si l'inégalité inverse de (2.90) se trouve vérifiée. L'utilisation de l'algorithme (2.90) est fondée sur l'hypothèse admissible que l'apparition d'un signal constant (positif) déplace à droite tous les quantiles de la distribution du bruit. Il y a lieu de noter que pour la réalisation de l'algorithme (2.90) on a besoin de $mn$ opérations au lieu de $n$ dans le cas du détecteur de signe.

## 2.4. ALGORITHMES DE RANG DE DÉTECTION DES SIGNAUX NOYÉS DANS DES BRUITS INDÉPENDANTS

**2.4.1. Détecteur signe-rang d'un signal constant.** Les détecteurs signe-rang des signaux sont en général plus efficaces que les détecteurs de signe que nous avons étudiés précédemment, étant également non paramétriques. Envisageons tout d'abord l'algorithme signe-rang de détection d'un signal constant (positif) noyé dans un bruit additif de moyenne nulle et de distribution symétrique. Cet algorithme non paramétrique, plus compliqué que l'algorithme de signe (2.47), utilise non seulement l'information concernant les signes des éléments de l'échantillon, mais également les rangs des valeurs absolues des observations.

Soient $\mathbf{x} = (x_1, \ldots, x_n)$ l'échantillon indépendant observé, et $R_i^+$ le rang de l'élément $|x_i|$. Un des algorithmes signe-rang possibles de détection d'un signal positif consiste à comparer avec

*) Le choix optimal des quantiles est étudié dans [16].

un seuil la somme des composantes du vecteur des rangs positifs, qui correspondent aux valeurs échantillonnées positives $x_i \geqslant 0$. On prend ainsi la décision que le signal est présent si

$$S_n(x) = \sum_{x_i \geqslant 0} R_i^+ = \sum_{i=1}^{n} R_i^+ u(x_i) \geqslant C, \tag{2.91}$$

où le seuil $C$ est déterminé avec une probabilité donnée $\alpha$ de fausse alarme *).

Etant donné que

$$\sum_{i=1}^{n} R_i^+ = \sum_{k=1}^{n} k = \frac{n(n+1)}{2},$$

on peut écrire comme suit l'inégalité (2.91) compte tenu de (2.39)

$$S_n(x) = \frac{1}{2} \sum_{i=1}^{n} R_i^+ \operatorname{sgn} x_i + \frac{n(n+1)}{4} \geqslant C. \tag{2.91'}$$

Le schéma de l'algorithme signe-rang envisagé comprend un redresseur, un dispositif de mise en rang des valeurs absolues des

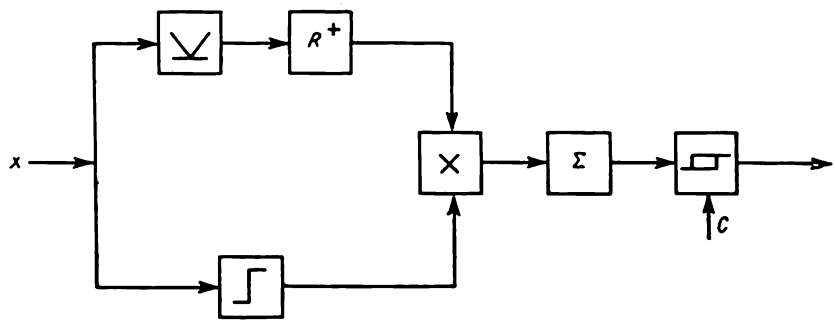



Fig. 2.3. Schéma fonctionnel du détecteur signe-rang d'un signal constant.

éléments de l'échantillon, un limiteur parfait, un multiplicateur, un sommateur et un dispositif de comparaison au seuil (fig. 2.3).

L'algorithme envisagé peut s'écrire sous une forme plus compacte en remarquant que $u(x_i + x_j) = 1$ si et seulement si $x_i \geqslant |x_j|$, $x_i > 0$ ou $x_j \geqslant |x_i|$, $x_j > 0$. Compte tenu de cette remarque on obtient

$$S_n(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} u(x_i + x_j) \geqslant C. \tag{2.92}$$

*) On appelle parfois l'algorithme (2.91) algorithme de rang de Wilcoxon à un échantillon (voir [5, 33]). Dans A.3.2. on trouvera une table empruntée dans [20] caractérisant la distribution d'une statistique signe-rang.

Pour un échantillon de taille finie il est assez difficile de trouver les fonctions de répartition de la statistique $S_n$ (x). Mais pour $n \to \infty$ en vertu du théorème de la limite centrale la distribution $S_n$ (x) est asymptotiquement normale. Nous allons trouver la valeur moyenne et la variance de la statistique $S_n$ (x) tant pour l'hypothèse $H$ (il n'y a pas de signal) que pour l'alternative $K$ (le signal est présent), ce qui nous permettra pour le cas $n \gg 1$ de déterminer, pour une probabilité donnée $\alpha$ de fausse alarme, le seuil $C$ dans l'algorithme (2.91) et la probabilité de perte de signal.

Notons tout d'abord que l'indépendance des vecteurs $R^+$ et sgn x pour un échantillon indépendant x d'un bruit stationnaire entraîne pour l'hypothèse $H$

$$m_1\{S_n(x)\} = \frac{1}{2}\sum_{i=1}^{n} m_1\{R_i^+\}\, m_1\{\operatorname{sgn} x_i\} + \frac{n(n+1)}{4},$$

et comme pour une distribution symétrique du bruit on a $m_1\{\operatorname{sgn} x_i\} = 0$, alors

$$m_1\{S_n(x)\} = n(n+1)/4 \tag{2.93}$$

et pour $n \gg 1$

$$m_1\{S_n(x)\} \sim n^2/4. \tag{2.93'}$$

Pour l'hypothèse $H$ la variance de la statistique $S_n$ (x) est égale à

$$M_2\{S_n(x)\} = \frac{1}{4} m_1\left\{\sum_{i=1}^{n}\sum_{j=1}^{n} R_i^+ R_j^+ \operatorname{sgn} x_i \operatorname{sgn} x_j\right\} = \frac{1}{4}\sum_{i=1}^{n} m_1\{R_i^{+2}\}.$$

Mais

$$m_1\{R_i^{+2}\} = \sum_{k=1}^{n} k^2 P\{R_i^+ = k\}$$

et pour un bruit stationnaire $P\{R_i^+ = k\} = \frac{1}{n}$. Donc

$$M_2\{S_n(x)\} = \frac{1}{4}\sum_{i=1}^{n}\left(\frac{1}{n}\sum_{k=1}^{n} k^2\right)$$

et comme $\sum_{k=1}^{n} k^2 = n(n+1)(2n+1)/6$, pour l'hypothèse $H$ on a

$$M_2\{S_n(x)\} = n(n+1)(2n+1)/24. \tag{2.94}$$

Pour $n \gg 1$

$$M_2\{S_n(x)\} \sim n^3/12. \tag{2.94'}$$

Pour trouver la valeur moyenne de la statistique $S_n$ (x) pour l'alternative $K$ servons-nous de la formule (2.92). Dans le second membre de cette formule on a $n$ termes pour $j = i$ et $n(n-1)/2$

pour $j \neq i$. La statistique $S_n(x)$ est égale au nombre des cas où $x_i + x_j \geqslant 0$. Par conséquent, pour l'alternative $K$, on a

$$m_1\{S_n(x)\} = nP\{x \geqslant 0 \mid K\} + \frac{n(n-1)}{2} P\{x+y>0 \mid K\}. \quad (2.95)$$

Mais

$$P\{x \geqslant 0 \mid K\} = 1 - F_1(0 \mid a), \quad (2.95')$$

et pour des $x$ et $y$ indépendants

$$P\{x+y>0 \mid K\} = 1 - \int_{-\infty}^{\infty} F_1(-z \mid a)\, w_1(z \mid a)\, dz, \quad (2.95'')$$

où $F_1(z \mid a)$ est la fonction de répartition et $w_1(z \mid a)$ la densité de probabilité de la somme d'un signal constant $a > 0$ et d'un bruit stationnaire. En utilisant (2.95) à (2.95″) on obtient

$$m_1\{S_n(x)\} = \frac{n(n+1)}{2} - nF_1(0 \mid a) - \\ - \frac{n(n-1)}{2} \int_{-\infty}^{\infty} F_1(-z \mid a)\, w_1(z \mid a)\, dz. \quad (2.96)$$

Pour une distribution symétrique du bruit additif $w_1(z \mid a) = w_1(z-a)$ pour $a \to 0$ on a

$$F_1(0 \mid a) = \int_{-\infty}^{0} w_1(z-a)\, dz = \frac{1}{2} + aw_1(0) + o(a), \quad (2.97)$$

$$\int_{-\infty}^{\infty} F_1(-z \mid a)\, w_1(z \mid a)\, dz = \int_{-\infty}^{\infty} F_1(-z-a)\, w_1(z-a)\, dz = \\ = \int_{-\infty}^{\infty} F_1(y-2a)\, w_1(-y)\, dy = \int_{-\infty}^{\infty} [F_1(y) - 2aw_1(y) + o(a)]\, w_1(y)\, dy = \\ = \frac{1}{2} - 2a \int_{-\infty}^{\infty} w_1^2(y)\, dy + o(a). \quad (2.97')$$

En substituant (2.97) et (2.97′) dans (2.96), on obtient pour $n \gg 1$ et a $\sqrt{n} = \text{const}$

$$m_1\{S_n(x)\} \sim \frac{n^2}{4} + n^2 a \int_{-\infty}^{\infty} w_1^2(y)\, dy. \quad (2.98)$$

D'une manière analogue on peut montrer que pour l'alternative $K$, lorsque $n \gg 1$ et $a\sqrt{n} = \text{const}$, on a

$$M_2\{S_n(x)\} \sim n^3/12. \quad (2.99)$$

En utilisant (2.93′) et (2.94′) et le fait que la statistique $S_n(x)$ est asymptotiquement normale, on obtient la probabilité de fausse alarme pour $n \gg 1$, soit

$$\alpha = 1 - F\left(\frac{C - n^2/4}{\sqrt{n^3/12}}\right),$$

d'où pour une valeur donnée de $\alpha$ on trouve le seuil $C$ dans l'algorithme (2.91), soit

$$C = \frac{n}{2}\left(x_\alpha \sqrt{\frac{n}{3}} + \frac{n}{2}\right), \tag{2.100}$$

où $x_\alpha$ est l'$\alpha$-quantile de la distribution normale.

En utilisant (2.98) et (2.99) on trouve la probabilité de perte de signal pour $n \gg 1$ et $a\sqrt{n} = \text{const}$

$$\beta = F\left[\left(C - \frac{n^2}{4} - n^2 a \int_{-\infty}^{\infty} w_1^2(y)\,dy\right) \Big/ \sqrt{n^3/12}\right]$$

ou, compte tenu de (2.100),

$$\beta = F\left(x_\alpha - \sqrt{12a^2 n} \int_{-\infty}^{\infty} w_1^2(y)\,dy\right). \tag{2.101}$$

On peut évidemment écrire la formule (2.101) sous la forme

$$x_{1-\beta} = x_\alpha - \sqrt{12a^2 n} \int_{-\infty}^{\infty} w_1^2(y)\,dy. \tag{2.101′}$$

**2.4.2. Efficacité relative du détecteur signe-rang.** Maintenant il n'est pas difficile de calculer le coefficient $\rho$ d'efficacité relative de l'algorithme signe-rang de détection d'un signal constant par rapport à l'algorithme linéaire (2.62). A cet effet il y a lieu de répéter le raisonnement de 2.3.3 et de remarquer que pour $a\sqrt{n} = \text{const}$ la formule (2.101′) peut être obtenue en remplaçant la grandeur $2w_1(0)$ par $\sqrt{12}\int_{-\infty}^{\infty} w_1^2(y)\,dy$ (voir également (2.68)).

On obtient alors

$$\rho = 12\sigma^2\left(\int_{-\infty}^{\infty} w_1^2(y)\,dy\right)^2, \tag{2.102}$$

où $w_1(y)$ est la densité de probabilité du bruit et $\sigma^2$ la variance du bruit.

Si la distribution du bruit est normale (de moyenne nulle) on a

$$\int_{-\infty}^{\infty} w_1^2(y)\,dy = \frac{1}{2\sigma\sqrt{\pi}}$$

et à partir de (2.102) on a $\rho = 3/\pi \approx 0{,}955$, c'est-à-dire que le détecteur signe-rang envisagé pour un bruit normal (pour un échantillon de taille importante et un signal faible) n'est qu'un point moins efficace que le détecteur optimal linéaire. Une augmentation considérable de l'efficacité par rapport à l'algorithme de signe (voir (2.71)) ne peut évidemment être atteinte qu'au prix d'une complication de l'algorithme. Alors que le nombre d'opérations pour l'algorithme de signe (2.47) croît linéairement avec la taille de l'échantillon, le nombre d'opérations dans un algorithme signe-rang est proportionnel au carré de la taille de l'échantillon.

Dans la table 2.1 ci-dessous sont indiquées les valeurs du coefficient d'efficacité asymptotique relative de l'algorithme signe-rang par rapport à l'algorithme linéaire pour les mêmes distributions du bruit que celles données dans 2.2.3.

La formule en bas de la table représente la distribution la moins favorable du bruit pour laquelle le coefficient $\rho$ prend la plus petite de ses valeurs *). Pour toutes les autres distributions $\rho > 0{,}864$.

*Table 2.1*

| Distribution du bruit | $\rho$ |
|---|---|
| Normale | 0,955 |
| Laplacienne | 1,5 |
| Uniforme | 1 |
| Sinusoïdale de phase aléatoire | $\infty$ |
| $w_1(x) = \frac{3}{4\sigma\sqrt{5}}\,[1 - x^2/(5\sigma^2)]$, $\lvert x \rvert \leqslant \sigma\sqrt{5}$ | 0,864 |

*) Il est facile de voir que le minimum de la fonctionnelle $\int_{-\infty}^{\infty} w_1^2(y)\,dy$ pour $\sigma^2$ donné et lorsque $w_1(y)$ est non négative et normée a lieu pour $w_1(y) = b - ay^2$, $\lvert y \rvert \leqslant \sqrt{b/a}$, $b > 0$, $a > 0$. Les constantes $a$ et $b$ sont données par les relations

$$2\int_0^{\sqrt{b/a}} x^2 w_1(y)\,dy = \sigma^2, \qquad 2\int_0^{\sqrt{b/a}} w_1(y)\,dy = 1.$$

En raisonnant comme à la fin de 2.3.6 on arrive à la conclusion que la formule (2.102) reste vraie pour le coefficient d'efficacité asymptotique relative de l'algorithme signe-rang envisagé de détection d'un signal constant, par rapport à l'algorithme (2.75) qui est optimal pour la détection d'un signal constant (positif) noyé dans un bruit normal additif de variance inconnue.

Avec les données de la table 2.1 et du paragraphe 2.3.6 et compte tenu de la formule (1.53') on peut facilement établir les valeurs du coefficient d'efficacité asymptotique relative de l'algorithme signe-rang (2.91) par rapport à l'algorithme de signe (2.47) (table 2.2).

Remarquons que pour la distribution laplacienne du bruit l'algorithme signe-rang est moins efficace que l'algorithme de signe.

Pour certaines classes de distributions il est possible d'augmenter l'efficacité du détecteur en utilisant des fonctions plus compliquées des statistiques signe-rang. Citons à titre d'exemple l'*algorithme signe-rang de Van der Waerden* en vertu duquel la décision sur la présence d'un signal constant (positif) dans un bruit additif à distribution symétrique est prise dans le cas où

$$\sum_{i=1}^{n} F^{-1}\left(\frac{1}{2}+\frac{1}{2}\frac{R_i^+}{n+1}\right) u(x_i) \geqslant C, \tag{2.103}$$

où $F^{-1}(x)$ est la fonction inverse de l'intégrale de Laplace.

*Table 2.2*

| Distribution du bruit | $\rho$ |
|---|---|
| Normale | 1,5 |
| Laplacienne | 0,75 |
| Uniforme | 3 |
| Sinusoïdale de phase aléatoire | $\infty$ |
| $w_1(x)=\frac{2}{4\sigma\sqrt{5}}[1-x^2/(5\sigma^2)],\ \lvert x\rvert \leqslant \sigma\sqrt{5}$ | 1,94 |

Jusqu'à présent nous avons envisagé des détecteurs signe-rang unilatéraux pour un signal constant a priori positif ($a > 0$). Il est évident que pour la détection d'un signal constant de signe quelconque ($a \neq 0$) il faut utiliser des algorithmes signe-rang bilatéraux à deux seuils. Il n'est pas nécessaire d'exposer la théorie de ces algorithmes car le lecteur obtiendra facilement les résultats correspondants par analogie avec la théorie des détecteurs de signe bilatéraux exposée ci-dessus.

**2.4.3. Algorithmes signe-rang mixtes de détection d'un signal.** L'augmentation de l'efficacité du détecteur signe-rang (2.92) par rapport à celle du détecteur de signe (2.47) est obtenue au prix d'un accroissement notable du nombre d'operations qui, dans l'algorithme (2.92), est proportionnel à $n^2$, et dans l'algorithme (2.47), à $n$. Il est possible cependant de diminuer notablement le nombre d'opérations du détecteur sans perdre en efficacité.

Supposons que l'on observe un échantillon indépendant de taille $n$. Formons $r$ groupes de valeurs échantillonnées à raison de $m$ valeurs par groupe ($mr = n$):

premier groupe: $x_1, x_2, \ldots, x_m$;
second groupe: $x_{m+1}, \ldots, x_{2m}$;
. . . . . .
groupe $i$: $x_{(i-1)m+1}, \ldots, x_{im}$;
. . . . . .
groupe $r$: $x_{(r-1)m+1}, \ldots, x_{rm}$.

Composons pour chacun de ces groupes d'éléments échantillonnés des statistiques signe-rang du type (2.92)

$$S_m^{(i)}(\mathbf{x}) = \sum_{j=1}^{m} \sum_{k=j}^{m} u(x_{im+j} + x_{im+k}), \quad j = 0, \ldots, r-1. \tag{2.104}$$

La décision sur la présence d'un signal constant noyé dans un bruit additif de distribution symétrique est prise si

$$\sum_{i=0}^{r-1} S_m^{(i)}(\mathbf{x}) \geqslant C. \tag{2.105}$$

L'algorithme de détection (2.105) est appelé *algorithme signe-rang mixte*. Le nombre d'opérations qu'entraîne le calcul de chacune des statistiques (2.104) étant proportionnel à $m^2$, le total des opérations de l'algorithme (2.105) est proportionnel à $rm^2 = nm$. Il est évident que pour $m \ll n$ le nombre d'opérations nécessaires à la réalisation de l'algorithme (2.105) sera $r$ fois inférieur à celui de l'algorithme (2.92). Le schéma de réalisation d'un algorithme signe-rang pour $m = 3$ est donné sur la figure 2.4. Bien qu'à première vue ce schéma soit plus compliqué que celui de la figure 2.3, on obtient, en réalité, une simplification grâce à un dispositif plus simple de mise en rang.

Les considérations suivantes permettent de supposer que l'efficacité de l'algorithme (2.105) est voisine de celle de l'algorithme (2.92) même pour $m$ relativement faible. En effet, pour des échantillons indépendants, pour $m \approx 10$ déjà la distribution de chacune des statistiques (2.104) diffère peu de la loi normale. Donc la sommation ultérieure des statistiques $S_m^{(i)}$, $i = 0, \ldots, r-1$, ne doit pas entraîner des pertes notables, car pour des statistiques normales le traitement linéaire est le meilleur.

Remarquons que pour $m = 1$ l'algorithme signe-rang mixte coïncide avec l'algorithme de signe, et pour $m \to n$ avec l'algorithme signe-rang.

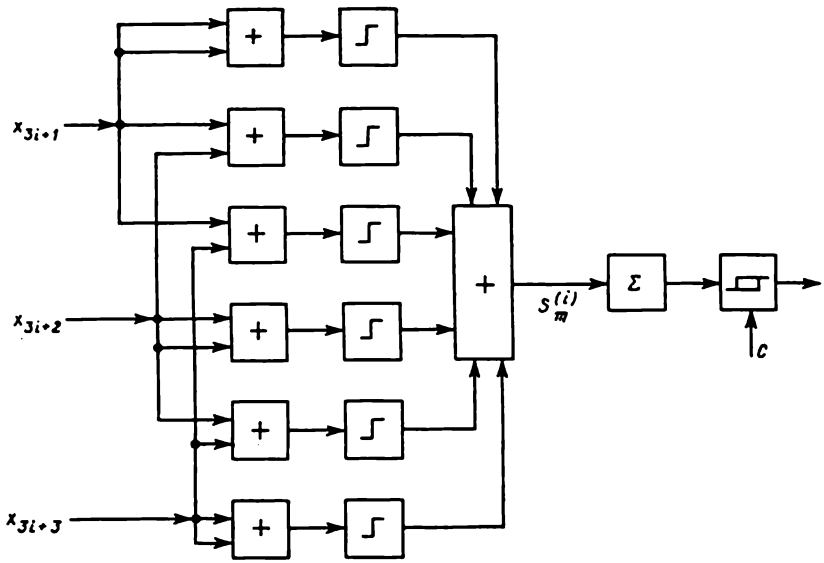



Fig. 2.4. Schéma fonctionnel du détecteur signe-rang mixte d'un signal constant.

**2.4.4. Efficacité relative de l'algorithme signe-rang mixte.** Ecrivons tout d'abord les expressions des valeurs moyennes des statistiques $S_m^{(i)}$ pour l'hypothèse $H$ et pour l'alternative $K$. En utilisant (2.93) et (2.95), (2.97) et (2.97'), on obtient

$$m_1\{S_m^{(i)}(\mathbf{x}) \mid H\} = m(m+1)/4, \tag{2.106}$$

$$m_1\{S_m^{(i)}(\mathbf{x}) \mid K\} = \frac{m^2}{2} + maw_1^2(0) + \\ + \frac{m(m-1)}{2}\left(\frac{1}{2} - 2a\int_{-\infty}^{\infty} w_1^2(z)\,dz\right) + o(a), \tag{2.107}$$

où $w_1(z)$ est la densité de probabilité du bruit.

La variance de la statistique $S_m^{(i)}(\mathbf{x})$ pour l'hypothèse $H$ est en vertu de (2.94) égale à

$$M_2\{S_m^{(i)}(\mathbf{x}) \mid H\} = m(m+1)(2m+1)/24. \tag{2.108}$$

(2.105) fournit immédiatement

$$m_1\left\{\sum_{i=0}^{r-1} S_m^{(i)}(x) \mid H\right\} = \frac{rm(m+1)}{4} = n(m+1)/4, \qquad (2.109)$$

$$m_1\left\{\sum_{i=0}^{r-1} S_m^{(i)}(x) \mid K\right\} = \frac{nm}{2} + naw_1(0) -$$

$$-\frac{n(m-1)}{2}\left(\frac{1}{2} - 2a\int_{-\infty}^{\infty} w_1^2(z)\,dz\right) + o(a) =$$

$$= \frac{n(m+1)}{4} + na\left[w_1(0) + (m-1)\int_{-\infty}^{\infty} w_1^2(z)\,dz\right] + o(a), \qquad (2.110)$$

$$M_2\left\{\sum_{i=0}^{r-1} S_m^{(i)}(x) \mid H\right\} = n(m+1)(2m+1)/24, \qquad (2.111)$$

d'où on obtient (voir (1.70'))

$$\varepsilon_1 = \frac{\left[w_1(0) + (m-1)\int_{-\infty}^{\infty} w_1^2(z)\,dz\right]^2}{(m+1)(2m+1)/24}. \qquad (2.112)$$

Maintenant il est facile d'obtenir l'expression du coefficient d'efficacité asymptotique relative de l'algorithme signe-rang mixte de détection d'un signal constant par rapport à l'algorithme linéaire (2.62) (voir (2.69'))

$$\rho = \frac{24\left[w_1(0) + (m-1)\int_{-\infty}^{\infty} w_1^2(z)\,dz\right]^2}{(m+1)(2m+1)}\,\sigma^2, \qquad (2.113)$$

où $\sigma^2$ est la variance du bruit.

La formule (2.69) de l'efficacité de l'algorithme de signe est un cas particulier de (2.113) pour $m = 1$. Pour $m \to \infty$ on obtient à partir de (2.113) la formule (2.102) pour l'algorithme signe-rang.

En vertu de (2.113) on a :

— pour un bruit à distribution normale

$$\rho = \frac{6(m-1+\sqrt{2})^2}{\pi(m+1)(2m+1)}, \qquad (2.114)$$

— pour un bruit à distribution de Laplace

$$\rho = \frac{3m+3}{2m+1}, \qquad (2.115)$$

— pour un bruit à distribution uniforme

$$\rho = \frac{2m^2}{(m+1)(2m+1)}. \qquad (2.116)$$

Dans la table comparative 2.3 des efficacités de l'algorithme mixte pour $m = 10$ et des algorithmes de signe et de signe-rang sont indiquées les valeurs des coefficients d'efficacité asymptotique relative de ces algorithmes par rapport à l'algorithme linéaire ainsi que l'ordre de grandeur du nombre d'opérations élémentaires pour chacun de ces algorithmes pour un échantillon de taille $n = 10^3$.

*Table 2.3*

| Type d'algorithme | ρ — Distribution | | | Nombre d'opérations |
|---|---|---|---|---|
| | normale | laplacienne | uniforme | |
| De signe | 0,64 | 2 | 0,33 | $10^3$ |
| Mixte ($m = 10$) | 0,91 | 1,57 | 0,88 | $10^4$ |
| Signe-rang | 0,95 | 1,5 | 1 | $10^6$ |

**2.4.5. Détecteurs signe-rang d'un signal déterministe quelconque.** Pour la détection d'un signal déterministe quelconque $\lambda s(t)$, $\lambda > 0$, noyé dans un bruit additif stationnaire de moyenne nulle et de distribution symétrique on peut utiliser une modification de l'algorithme (2.91). Notamment, on prend la décision sur la présence du signal dans le cas où

$$S_n(\mathbf{x}) = \sum_{i=1}^{n} s_i R_i^+ u(x_i) \geqslant C, \tag{2.117}$$

où $s_i = s(t_i)$, $i = 1, \ldots, n$. L'inégalité (2.117) peut également s'écrire sous la forme (voir (2.91′) et (2.92))

$$S_n(\mathbf{x}) = \frac{1}{2} \sum_{i=1}^{n} s_i R_i^+ \operatorname{sgn} x_i + \frac{1}{2} \sum_{i=1}^{n} s_i R_i^+ \geqslant C \tag{2.118}$$

ou

$$S_n(\mathbf{x}) = \sum_{i=1}^{n} s_i \sum_{j=i}^{n} u(x_i + x_j) \geqslant C. \tag{2.118′}$$

En vertu du théorème de la limite centrale la statistique $S_n(\mathbf{x})$ dans le cas envisagé est asymptotiquement normale pour $n \to \infty$ à condition que

$$\lim_{n\to\infty} \max_{1\leqslant i\leqslant n} s^2(t_i) / \sum_{i=1}^{n} s_2(t_i) = 0. \tag{2.119}$$

Pour les signaux utilisés dans la pratique cette condition est toujours vérifiée. Calculons la valeur moyenne et la variance de la statistique $S_n(\mathbf{x})$.

Pour l'hypothèse $H$ (il n'y a pas de signal) on obtient à partir de (2.118)

$$m_1\{S_n(\mathbf{x})\mid H\}=\frac{1}{2}\sum_{i=1}^{n}s_i m_1\{R_i^+\}=\frac{1}{2}\sum_{i=1}^{n}s_i\sum_{k=1}^{n}kP\{R_i^+=k\}=$$

$$=\frac{1}{2}\sum_{i=1}^{n}s_i\left(\frac{1}{n}\sum_{k=1}^{n}k\right)=\frac{n+1}{4}\sum_{i=1}^{n}s_i$$

ou

$$m_1\{S_n(\mathbf{x})\mid H\}=\frac{n(n+1)}{4}a_{sn}, \tag{2.120}$$

où

$$a_{sn}=\frac{1}{n}\sum_{i=1}^{n}s(t_i) \tag{2.120'}$$

est la composante constante du signal.

Pour $n\gg 1$ on a en vertu de (2.120)

$$m_1\{S_n(\mathbf{x})\mid H\}\sim\frac{n^2}{4}a_{sn}. \tag{2.121}$$

Le carré moyen de la statistique $S_n(\mathbf{x})$ pour l'hypothèse $H$ est égal à (voir [5])

$$m_1\{S_n^2(\mathbf{x})\mid H\}=\frac{1}{4}m_1\Big\{\sum_{i=1}^{n}\sum_{j=1}^{n}s_is_j(R_i^+R_j^+\operatorname{sgn}x_i\operatorname{sgn}x_j+$$

$$+R_i^+R_j^+\operatorname{sgn}x_i+R_i^+R_j^+\operatorname{sgn}x_j+R_i^+R_j^+)\Big\}=$$

$$=\frac{1}{4}\sum_{i=1}^{n}s_i^2m_1\{R_i^{+2}\}+\frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{n}s_is_jm_1\{R_i^+R_j^+\}=$$

$$=\frac{1}{4}\sum_{i=1}^{n}s_i^2\left(\frac{1}{n}\sum_{k=1}^{n}k^2\right)+\frac{1}{4}\left(\sum_{i=1}^{n}s_im_1\{R_i^+\}\right)^2,$$

d'où l'on tire la variance $M_2$ de la statistique $S_n(\mathbf{x})$ pour l'hypothèse $H$

$$M_2\{S_n(\mathbf{x})\mid H\}=m_1\{S_n^2(\mathbf{x})\mid H\}-[m_1\{S_n(\mathbf{x})\mid H\}]^2=$$

$$=\frac{(n+1)(2n+1)}{24}\sum_{i=1}^{n}s_i^2$$

ou

$$M_2\{S_n(\mathbf{x})\mid H\}=\frac{n(n+1)(2n+1)}{24}W_{sn}, \tag{2.122}$$

où

$$W_{sn} = \frac{1}{n} \sum_{i=1}^{n} s^2(t_i). \tag{2.122'}$$

Pour $n \gg 1$ (2.122) entraîne

$$M_2\{S_n(x) \mid H\} \sim \frac{n^3}{12} W_{sn}. \tag{2.123}$$

En utilisant (2.121), (2.123) et le fait que la statistique $S_n(x)$ est asymptotiquement normale, on trouve la valeur du seuil $C$ dans l'algorithme (2.117) pour $n \gg 1$ (voir (2.100))

$$C = \frac{n\sqrt{W_{sn}}}{2} \left( x_\alpha \sqrt{\frac{n}{3}} + \frac{n}{2} \frac{a_{sn}}{\sqrt{W_{sn}}} \right), \tag{2.124}$$

où $x_\alpha$ est l'$\alpha$-quantile de la loi normale.

Pour trouver la valeur moyenne de la statistique $S_n(x)$ pour l'alternative $K$ on se réfère à la formule (2.118'). Tout comme pour (2.95), on a

$$m_1\{S_n(x) \mid K\} = \sum_{i=1}^{n} s_i P\{x_i \geqslant 0 \mid K\} + \sum_{i=1}^{n} \sum_{j=i+1}^{n} s_i P\{x_i + x_j > 0 \mid K\}. \tag{2.125}$$

Et d'une manière analogue à (2.97) et (2.97'), dans le cas envisagé on a

$$P\{x_i \geqslant 0 \mid K\} = \frac{1}{2} + \lambda s_i w_1(0) + o(\lambda), \tag{2.126}$$

$$P\{x_i + x_j > 0 \mid K\} = 1 - \int_{-\infty}^{\infty} F_1(y - s_i - s_j)\, w_1(y)\, dy + o(\lambda) =$$

$$= \frac{1}{2} + \lambda(s_i + s_j) \int_{-\infty}^{\infty} w_1^2(y)\, dy + o(\lambda). \tag{2.126'}$$

En substituant (2.126), (2.126') dans (2.125), on obtient pour l'alternative $K$

$$m_1\{S_n(x) \mid K\} = \frac{n}{2} a_{sn} - n w_1(0) \lambda W_{sn} + \frac{n(n-1)}{4} a_{sn} +$$

$$+ \lambda^2 n^2 \int_{-\infty}^{\infty} w_1^2(y)\, dy \left[ \frac{1}{n} \sum_{i=1}^{n} \left(1 - \frac{i+1}{n}\right) s_i^2 + \right.$$

$$\left. + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=i+1}^{n} s_i s_j \right] + o(\lambda). \tag{2.127}$$

Etant donné que

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \left(1 - \frac{i+1}{n}\right) s_i^2 = W_s, \quad W_s = \lim_{n\to\infty} W_{sn},$$

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=i+1}^{n} s_i s_j = a_s^2 - W_s, \quad a_s = \lim_{n\to\infty} a_{sn},$$

on obtient à partir de (2.127) pour $n \gg 1$ et $\lambda\sqrt{n} = \text{const}$

$$m_1\{S_n(\mathbf{x}) \mid K\} \sim \frac{n^2 a_s}{4} + n^2\lambda a_s^2 \int_{-\infty}^{\infty} w_1^2(y)\, dy. \tag{2.128}$$

La probabilité de perte du signal pour $n \gg 1$ est égale à

$$\beta = F\left(x_\alpha - \sqrt{\frac{12\lambda^2 n}{W_s}}\, a_s^2 \int_{-\infty}^{\infty} w_1^2(y)\, dy\right) \tag{2.129}$$

ou

$$x_{1-\beta} = x_\alpha - \sqrt{\frac{12\lambda^2 n}{W_s}}\, a_s^2 \int_{-\infty}^{\infty} w_1^2(y)\, dy. \tag{2.129'}$$

En remarquant que (2.129′) ne diffère de (2.101′) que par le facteur $a_s^2/\sqrt{W_s}$ dans le second terme du membre de droite, on trouve avec (2.87) et (2.102) l'expression suivante pour le coefficient d'efficacité asymptotique relative de l'algorithme signe-rang (2.117) par rapport à l'algorithme linéaire (2.86)

$$\rho = 12\sigma^2 \frac{a_s^4}{W_s} \left(\int_{-\infty}^{\infty} w_1^2(y)\, dy\right)^2. \tag{2.130}$$

Pour $a_s = 0$, i.e. lorsque le signal n'a pas de composante constante, le coefficient d'efficacité asymptotique relative de l'algorithme signe-rang s'annule.

Pour un signal constant $s(t) \equiv 1$ le rapport $a_s^4/W_s = 1$, et la formule (2.130) coïncide avec (2.102). Pour une suite périodique d'impulsions non modulées de durée $\tau$ et de période $T$ ce rapport est égal à $(\tau/T)^3$ et pour une modulation d'amplitude de taux $m_a$ on a

$$a_s^4/W_s = (\tau/T)^3 (1 + m_a^2/2)^{-1}.$$

**2.4.6. Algorithmes de rang de détection des signaux à bande étroite.** Le fait que les signaux haute fréquence à bande étroite ne contiennent pas de composante constante rend l'utilisation des algorithmes signe-rang pour un traitement prédétecteur des observations absolument inefficace, car par rapport à l'algorithme linéaire le

coefficient d'efficacité asymptotique relative de ces algorithmes, comme nous l'avons vu plus haut, est nul pour une composante constante nulle. Les algorithmes de rang de détection non paramétrique des signaux radio-électriques à bande étroite (sans composante constante) sont plus efficaces. Les algorithmes de rang peuvent évidemment être utilisés également pour la détection des signaux contenant une composante constante. Cependant, comme nous allons le montrer plus loin, ils sont moins efficaces que les algorithmes signe-rang correspondants.

Soient $\lambda s(t)$, $\lambda > 0$, un signal déterministe, et $\mathbf{x} = (x_1, \ldots, x_n)$ l'échantillon indépendant d'observations appartenant soit à un bruit

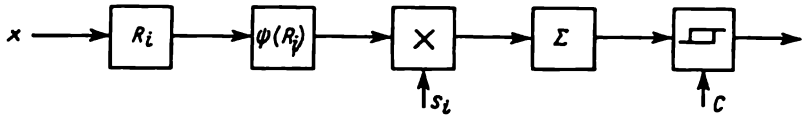



Fig. 2.5. Schéma fonctionnel du détecteur de rang d'un signal déterministe.

stationnaire (dont la densité de probabilité n'est pas obligatoirement symétrique), soit à la somme d'un signal déterministe et d'un bruit. La règle de choix d'une décision, basée sur la *statistique linéaire de rang*, se formule comme suit: le signal est présent si

$$T_n(\mathbf{x}) = \sum_{i=1}^{n} s_i \psi(R_i) \geqslant C, \tag{2.131}$$

où $s_i = s(t_i)$; $R_i$ est le rang de $x_i$, $i = 1, \ldots, n$, et $\psi(k)$ une certaine fonction de l'argument entier $k = 1, \ldots, n$. Le schéma de l'algorithme de rang utilisé pour la détection d'un signal déterministe est donné sur la figure 2.5.

Les exemples suivants sont des cas particuliers de l'algorithme général (2.131):

l'algorithme à médiane

$$\sum_{i=1}^{n} s_i \operatorname{sgn}\left(R_i - \frac{n+1}{2}\right) \geqslant C; \tag{2.132}$$

l'algorithme de Wilcoxon *)

$$\sum_{i=1}^{n} s_i R_i \geqslant C; \tag{2.133}$$

*) L'algorithme de Wilcoxon à échantillon double envisagé dans 2.1.7 est un cas particulier de (2.133) pour

$$s_i = 1;\ i \leqslant m,\ s_i = 0,\ i = m+1, \ldots, n.$$

l'algorithme de Van der Waerden

$$\sum_{i=1}^{n} s_i F^{-1}\left(\frac{R_i}{n+1}\right) \geqslant C, \tag{2.134}$$

où $F^{-1}(x)$ est la fonction inverse de l'intégrale de Laplace.

Déterminons la valeur moyenne et la variance de la statistique $T_n(\mathbf{x})$ pour l'hypothèse $H$ (il n'y a pas de signal). Comme pour un bruit stationnaire la distribution des rangs est uniforme, on a pour l'hypothèse $H$

$$m_1\{T_n(\mathbf{x}) \mid H\} = \sum_{i=1}^{n} s_i m_1\{\psi(R_i)\} = \frac{1}{n}\sum_{i=1}^{n} s_i \sum_{k=1}^{n} \psi(k)$$

ou

$$m_1\{T_n(\mathbf{x}) \mid H\} = n a_{sn} \overline{\psi}_n, \tag{2.135}$$

où

$$a_{sn} = \frac{1}{n}\sum_{i=1}^{n} s_i, \quad \overline{\psi}_n = \frac{1}{n}\sum_{k=1}^{n} \psi(k). \tag{2.136}$$

La variance de la statistique $T_n$ pour l'hypothèse $H$ est

$$M_2\{T_n(\mathbf{x}) \mid H\} = m_1\{T_n^2(\mathbf{x}) \mid H\} - [m_1\{T_n(\mathbf{x}) \mid H\}]^2 = n b_{sn}^2 \overline{\psi}_n^2, \tag{2.137}$$

où

$$b_{sn}^2 = \frac{1}{n}\sum_{i=1}^{n} (s_i - a_s)^2 = W_{sn} - a_{sn}^2, \tag{2.138}$$

$$\overline{\psi}_n^2 = \frac{1}{n-1}\sum_{k=1}^{n} [\psi(k) - \overline{\psi}_n]^2. \tag{2.138'}$$

Si $a_{sn} = 0$, on a pour l'hypothèse $H$

$$m_1\{T_n(\mathbf{x}) \mid H\} = 0, \quad M_2\{T_n(\mathbf{x}) \mid H\} = n W_{sn} \overline{\psi}_n^2. \tag{2.139}$$

Si la condition (2.119) est vérifiée, la statistique $T_n$ est asymptotiquement normale et le seuil $C$ dans l'algorithme (2.131) est lié à la probabilité de fausse alarme $\alpha$ par la relation suivante pour $n \gg 1$

$$C = x_\alpha (n b_{sn}^2 \overline{\psi}_n^2)^{1/2} + n a_{sn} \overline{\psi}_n. \tag{2.140}$$

**2.4.7. Coefficient d'efficacité asymptotique relative de l'algorithme de Wilcoxon.** Le calcul de la valeur moyenne et de la variance de la statistique $T_n$ pour l'alternative $K$ avec une fonction $\psi$ quel-

conque est un problème ardu *). Nous allons nous limiter à l'algorithme linéaire de rang de Wilcoxon (2.133); écrivons-le sous la forme (voir (2.41))

$$T_n(\mathbf{x}) = \sum_{i=1}^{n} s_i \sum_{j=1}^{n} u(x_i - x_j) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} s_i \operatorname{sgn}(x_i - x_j) + \frac{n^2 a_s}{2} \geqslant C.$$

Calculons la valeur moyenne de sgn $(x_i - x_j)$ pour l'alternative $K$:

$$m_1\{\operatorname{sgn}(x_i - x_j) \mid K\} = 1 - 2P\{x_i - x_j < 0 \mid K\} =$$

$$= 1 - 2\int_{-\infty}^{\infty} F_1(z - \lambda s_i)\, w_1(z - \lambda s_j)\, dz =$$

$$= 1 - 2\int_{-\infty}^{\infty} F_1(z + \lambda s_j - \lambda s_i)\, w_1(z)\, dz, \qquad (2.141)$$

où $w_1(z)$, $F_1(z)$ sont respectivement la densité de probabilité (non obligatoirement symétrique) et la fonction de répartition d'un bruit stationnaire additif. En développant $F_1(z + \lambda s)$ en une série suivant les puissances de $\lambda$, on obtient:

$$m_1\{\operatorname{sgn}(x_i - x_j) \mid K\} = 2\lambda(s_i - s_j)\int_{-\infty}^{\infty} w_1^2(z)\, dz + o(\lambda). \qquad (2.142)$$

(2.142) permet de trouver la valeur moyenne de la statistique $T_n$ pour l'alternative $K$ ($n \gg 1$, $\lambda\sqrt{n} = \text{const}$), soit

$$m_1\{T_n(\mathbf{x}) \mid K\} = \lambda \int_{-\infty}^{\infty} w_1^2(z)\, dz \sum_{i=1}^{n} \sum_{j=1}^{n} s_i(s_i - s_j) + \frac{n^2 a_{sn}}{2} =$$

$$= \lambda n^2 \int_{-\infty}^{\infty} w_1^2(z)\, dz\,(W_{sn} - a_{sn}^2) + \frac{n^2 a_{sn}}{2}. \qquad (2.143)$$

Ecrivons la probabilité de perte de signal avec l'algorithme (2.133) **), ayant en vue que la statistique $T_n$ est asymptotiquement

*) A cet effet on peut utiliser le théorème de Tchernov-Savage (voir [5, 21, 22]).

**) Il est facile de montrer que pour $n \gg 1$, $\lambda\sqrt{n} = \text{const}$

$$M_2\{T_n(\mathbf{x}) \mid H\} = M_2\{T_n(\mathbf{x}) \mid K\} \sim n^3 b_{sn}^2/12$$

(voir (2.137) pour $\psi(k) = k$).

normale pour $n \gg 1$:

$$\beta \sim F\left[\frac{C-\frac{n^2 a_{sn}}{2}-\lambda\left(W_{sn}-a_{sn}^2\right) n^2 \int\limits_{-\infty}^{\infty} w_1^2(z)\, dz}{b_{sn}\sqrt{n^3/12}}\right] \tag{2.144}$$

et comme dans le cas envisagé en vertu de (2.140) on a

$$C = x_\alpha b_{sn}\sqrt{n^3/12} + n^2 a_{sn}/2, \tag{2.145}$$

donc

$$x_{1-\beta} = x_\alpha - \lambda b_{sn}\sqrt{12n}\int\limits_{-\infty}^{\infty} w_1^2(z)\, dz. \tag{2.146}$$

Il n'est pas difficile d'écrire maintenant le coefficient d'efficacité asymptotique relative de l'algorithme de rang de Wilcoxon (2.133) par rapport à l'algorithme linéaire (2.86)

$$\rho = \frac{12\sigma^2 b_s^2}{W_s}\left(\int\limits_{-\infty}^{\infty} w_1^2(z)\, dz\right)^2, \tag{2.147}$$

où $\sigma^2$ est la variance du bruit $b_s = \lim\limits_{n\to\infty} b_{sn}$.

Si la composante constante du signal est nulle, on a $b_s^2 = W_s$ et à partir de (2.147) on trouve

$$\rho = 12\sigma^2\left(\int\limits_{-\infty}^{\infty} w_1^2(z)\, dz\right)^2. \tag{2.148}$$

Le rapport du coefficient (2.147) au coefficient (2.148) donne le coefficient d'efficacité asymptotique relative de l'algorithme de rang (2.133) pour la détection du signal $s(t)$ par rapport à l'algorithme (2.133) utilisé pour la détection du même signal centré ($a_s = 0$):

$$\rho = b_s^2/W_s = 1 - a_s^2/W_s. \tag{2.148'}$$

Par conséquent, l'efficacité de l'algorithme de rang (2.133) utilisé pour la détection d'un signal non centré diminue par rapport au cas d'un signal centré d'une grandeur égale au rapport du carré de la composante constante à la puissance du signal. Il est évident que $a_s^2 \leqslant W_s$ (car $b_s^2 \geqslant 0$, avec $b_s = 0$ pour un signal constant $s(t) \equiv a$).

Comparons enfin l'algorithme de rang (2.133) avec l'algorithme signe-rang (2.117). En vertu de (2.130) et (2.147), le coefficient d'efficacité asymptotique relative de l'algorithme (2.133) par rapport à l'algorithme (2.117) est égal à

$$\rho = b_s^2 W_s/a_s^4 = \frac{W_s}{a_s^2}\left(\frac{W_s}{a_s^2}-1\right). \tag{2.148''}$$

En vertu de (2.148″) on a $\rho \to \infty$ pour $a_s \to 0$, ce qui correspond à l'efficacité nulle signalée plus haut de l'algorithme signe-rang lors de la détection d'un signal sans composante constante. Au contraire, pour $a_s^2 \to W_s$ l'efficacité de l'algorithme de rang par rapport à l'alorithme signe-rang tend vers zéro. Pour la valeur limite du rapport $a_s^2/W_s = 0{,}62$ on a (voir (2.148″)) $\rho = 1$, i.e. ces deux algorithmes ont même efficacité.

Ainsi, pour $a_s^2 W_s < 0{,}62$ la détection d'un signal déterministe est plus sûre avec un algorithme de rang, et pour $a_s^2/W_s > 0{,}62$ c'est un algorithme signe-rang qui est plus efficace. Cependant, dans le cas d'un signal centré, l'algorithme de Wilcoxon

$$\sum_{i=1}^{n} (s_i - a_s)\, R_i \geqslant C \tag{2.149}$$

est préférable à l'algorithme signe-rang (2.117), le coefficient d'efficacité asymptotique relative du premier par rapport au second calculé à l'aide de (2.130) et (2.148) étant égal à $(W_s/a_s^2) \geqslant 1$.

**2.4.8. Algorithme de rang simplifié de détection des signaux.** Comme nous l'avons noté dans 2.4.3, pour réduire le nombre d'opérations liées à la mise en rang d'un échantillon de taille importante

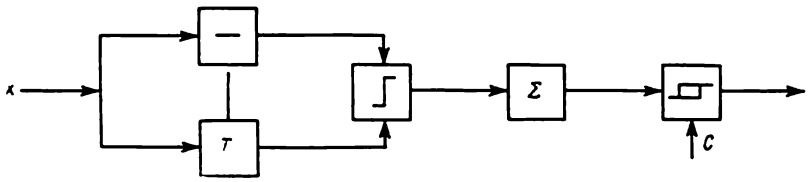



Fig. 2.6. Schéma fonctionnel d'un détecteur de rang simple.

on peut utiliser les statistiques mixtes, en divisant l'échantillon initial en groupes et en procédant à la mise en rang dans chacun des groupes. Cet artifice peut également être appliqué aux algorithmes de rang du paragraphe précédent.

Voyons encore une méthode de réduction du nombre d'opérations dans un algorithme de rang non paramétrique sur l'exemple de l'algorithme de Wilcoxon (2.141). Un algorithme de rang très simple de détection d'un signal déterministe proposé dans [13] impose la comparaison avec un seuil de la somme

$$\sum_{i=1}^{n-1} s_{i+1} \operatorname{sgn}(x_{i+1} - x_i) \geqslant C_1. \tag{2.150}$$

L'algorithme (2.150) est schématisé par la figure 2.6. Le dispositif réalisant cet algorithme est obtenu à partir du système usuel

limiteur parfait — filtre à l'entrée duquel on amène non pas les observations mais la différence des deux lectures successives. Il est clair qu'une telle simplification entraîne une diminution de l'efficacité de l'algorithme (2.150) par rapport à (2.141).

Pour l'hypothèse $H$ (il n'y a pas de signal), en opérant sur un échantillon indépendant appartenant à la distribution d'un bruit stationnaire, on trouve

$$m_1\left\{\sum_{i=1}^{n-1} s_{i+1}\operatorname{sgn}(x_{i+1}-x_i)\right\}=\sum_{i=1}^{n-1} s_{i+1}m_1\{\operatorname{sgn}(x_{i+1}-x_i)\}=0, \quad (2.151)$$

$$M_2\left\{\sum_{i=1}^{n-1} s_{i+1}\operatorname{sgn}(x_{i+1}-x_i)\right\}=\sum_{i=1}^{n-1}\sum_{j=1}^{n-1} s_{i+1}s_{j+1}\times$$

$$\times m_1\{\operatorname{sgn}(x_{i+1}-x_i)\operatorname{sgn}(x_{j+1}-x_j)\}=$$

$$=\sum_{i=1}^{n-1} s_{i+1}^2+2\sum_{i=1}^{n-1} s_{i+1}s_{i+2}m_1\{\operatorname{sgn}(x_{i+2}-x_{i+1})\operatorname{sgn}(x_{i+1}-x_i)\}. \quad (2.151')$$

Pour l'alternative $K$ (le signal est présent), en utilisant (2.142) on obtient pour $n \gg 1$, $\lambda\sqrt{n}=\text{const}$

$$m_1\left\{\sum_{i=1}^{n-1} s_{i+1}\operatorname{sgn}(x_{i+1}-x_i)\right\}\sim 2\lambda\int_{-\infty}^{\infty} w_1^2(z)\,dz\sum_{i=1}^{n-1} s_{i+1}(s_{i+1}-s_i). \quad (2.152)$$

Lorsque la condition (2.119) est vérifiée, la statistique dans (2.150) est asymptotiquement normale. Connaissant la moyenne et la variance pour l'hypothèse et pour l'alternative (formules (2.151), (2.151'), (2.152)), on peut estimer la perte en efficacité de l'algorithme (2.150) par rapport à (2.141) lors de la détection d'un signal déterministe (à composante constante nulle). On se convainc aisément que le coefficient d'efficacité asymptotique relative de l'algorithme de rang simplifié (2.150) par rapport à l'algorithme de rang (2.141) (voir (2.148)) est égal à

$$\rho=\frac{(1+R_s)^2}{3(1-R_{sx})}, \quad (2.153)$$

où

$$R_s=\frac{1}{W_s}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n-1} s_{i+1}s_i=\frac{1}{W_s}\lim_{T\to\infty}\frac{1}{T}\int_0^T s(t+\Delta)\,s(t)\,dt, \quad (2.153')$$

$$R_{sx}=\frac{2}{W_s}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n-1} s_{i+1}s_{i+2}m_1\{\operatorname{sgn}(x_{i+1}-x_{i+2})\operatorname{sgn}(x_{i+1}-x_i)\}=$$

$$=\frac{1}{W_s}\lim_{T\to\infty}\frac{1}{T}\int_0^T s(t+\Delta)\,s(t+2\Delta)\,m_1\{\operatorname{sgn}[x(t+\Delta)-x(t+2\Delta)]\times$$

$$\times\operatorname{sgn}[x(t+\Delta)-x(t)]\}\,dt, \quad (2.153'')$$

$\Delta$ étant l'intervalle de quantification des observations dans le temps; $x(t)$ est une réalisation du bruit.

Dans le cas où les algorithmes de rang (2.141) et (2.150) sont utilisés pour la détection d'un signal noyé dans un bruit normal additif de variance $\sigma^2$, on obtient $R_{sx} = {}^2/_3 R_s$, car

$$M_2\{x_{i+1} - x_i\} = 2\sigma^2,$$

$$r = m_1\{(x_{i+1} - x_{i+2})(x_{i+1} - x_i)\}/2\sigma^2 = m_1\{x_{i+1}^2\}/2\sigma^2 = \frac{1}{2},$$

et donc

$$m_1\{\operatorname{sgn}(x_{i+1} - x_{i+2}) \operatorname{sgn}(x_{i+1} - x_i)\} = \frac{2}{\pi} \arcsin \frac{1}{2} = \frac{1}{3}.$$

Ainsi, dans le cas d'un bruit normal additif le coefficient d'efficacité asymptotique relative de l'algorithme (2.150) par rapport à l'algorithme (2.141) est donné par

$$\rho = (1 - R_s)^2 \Big/ \left[3\left(1 - \frac{2}{3} R_s\right)\right], \quad |R_s| \leqslant 1, \tag{2.154}$$

par conséquent, $\rho \leqslant 4/5$, c'est-à-dire que la perte en efficacité est toujours supérieure à 20 %. La valeur maximale du coefficient $\rho$ correspond à $R_s = -1$ et la valeur minimale $\rho = 0$ à $R_s = 1$.

On conçoit bien qu'un algorithme de rang à canaux multiples utilisant les éléments de $k \leqslant n - 1$ diagonales de la matrice $[\operatorname{sgn}(x_i - x_j)]$ au lieu d'une est plus efficace. La statistique résultante s'obtient par sommation des statistiques de tous les canaux. Pour $k = n - 1$ on obtient l'algorithme utilisant les statistiques mixtes qu'on forme à partir de tous les éléments de l'échantillon des observations.

## 2.5. ALGORITHMES NON PARAMÉTRIQUES DE DÉTECTION DES SIGNAUX STOCHASTIQUES NOYÉS DANS DES BRUITS INDÉPENDANTS

**2.5.1. Système optimal à deux canaux de détection d'un signal normal noyé dans un bruit normal.** Pour la détection d'un signal stochastique (fluctuant) noyé dans des bruits additifs indépendants on utilise parfois un système à deux canaux (réception en diversité). En l'absence de signal, dans chacun des canaux seuls les bruits sont présents et les processus observés sont indépendants. Lorsque les signaux apparaissent dans les deux canaux, une corrélation statistique est observée entre ces processus aléatoires.

Considérons tout d'abord le problème de la détection d'un signal normal noyé dans des bruits normaux additifs. Supposons que les valeurs moyennes du signal et du bruit sont nulles et que les variances du signal et du bruit sont données et égales respectivement

à $\sigma_s^2$ et $\sigma^2$. Soient $x(t)$ et $y(t)$ les processus observés dans le premier et le second canal; en l'absence de signal (hypothèse $H$) on a

$$M_2\{x(t)\} = M_2\{y(t)\} = \sigma^2, \tag{2.155}$$

$$R_{xy} = m_1\{x(t)\,y(t)\}/\sigma^2 = 0 \tag{2.155'}$$

et si le signal est présent (alternative $K$)

$$M_2\{x(t)\} = M_2\{y(t)\} = \sigma^2 + \sigma_s^2, \tag{2.156}$$

$$R_{xy} = \frac{m_1\{x(t)\,y(t)\}}{\sigma^2+\sigma_s^2} = \frac{\sigma_s^2}{\sigma^2+\sigma_s^2} = r. \tag{2.156'}$$

Le logarithme du rapport de vraisemblance pour une observation est égal à

$$\ln l(x, y) = \ln\frac{w_2(x, y \mid K)}{w_2(x, y \mid H)} = -\frac{x^2-2rxy+y^2}{2(\sigma^2+\sigma_s^2)(1-r^2)} +$$

$$+\frac{x^2+y^2}{2\sigma^2} - \frac{1}{2}\ln\left[(1-r^2)\frac{\sigma^2+\sigma_s^2}{\sigma^2}\right] = \frac{\sigma_s^2(x+y)^2}{2[(\sigma^2+\sigma_s^2)^2-\sigma_s^4]} -$$

$$-\frac{1}{2}\ln[(1-r^2)(\sigma^2+\sigma_s^2)/\sigma^2].$$

Pour $n$ observations indépendantes dans chacun des canaux on a

$$\ln l(x_1, \ldots, x_n, y_1, \ldots, y_n) =$$

$$= \frac{\sigma_s^2}{2[(\sigma^2+\sigma_s^2)^2-\sigma_s^4]}\sum_{i=1}^{n}(x_i+y_i)^2 - \frac{n}{2}\ln[(1-r^2)(\sigma^2+\sigma_s^2)/\sigma^2]. \tag{2.157}$$

La relation (2.157) donne la règle optimale d'après le critère de Neyman-Pearson de détection du signal: le signal est présent si

$$\sum_{i=1}^{n}(x_i+y_i)^2 \geqslant C, \tag{2.158}$$

et il n'y a pas de signal si c'est l'inégalité inverse de (2.158) qui est vérifiée. Le schéma du détecteur réalisant l'algorithme (2.158) est donné sur la figure 2.7.

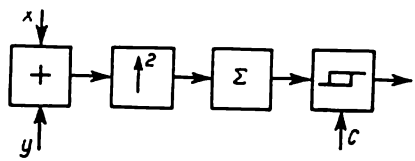



Fig. 2.7. Détecteur optimal à deux canaux d'un signal aléatoire normal noyé dans un bruit normal additif.

Pour l'hypothèse $H$ la somme des grandeurs aléatoires dans (2.158) est asymptotiquement normale de paramètres

$$m_1 \{ \sum_{i=1}^{n} (x_i + y_i)^2 \mid H \} = 2n\sigma^2, \tag{2.159}$$

$$M_2 \{ \sum_{i=1}^{n} (x_i + y_i)^2 \mid H \} = 8n\sigma^4. \tag{2.159'}$$

Pour $n \gg 1$ et une probabilité donnée de fausse alarme $\alpha$, on trouve à partir de (2.159), (2.159') la valeur du seuil $C$

$$C = \sigma^2 \sqrt{8n}\, (x_\alpha + \sqrt{n/2}). \tag{2.160}$$

Pour l'alternative $K$ la somme dans (2.158) est également asymptotiquement normale, avec

$$m_1 \{ \sum_{i=1}^{n} (x_i + y_i)^2 \mid K \} = 2n\,(\sigma^2 + 2\sigma_s^2), \tag{2.161}$$

et pour un *signal faible* ($\sigma_s^2 \ll \sigma^2$) on a

$$M_2 \{ \sum_{i=1}^{n} (x_i + y_i)^2 \mid K \} \sim M_2 \{ \sum_{i=1}^{n} (x_i + y_i)^2 \mid H \}. \tag{2.161'}$$

Compte tenu de (2.160) on trouve la probabilité de perte du signal

$$\beta = F\left(x_\alpha - \sqrt{2n}\, \sigma_s^2/\sigma^2\right) \tag{2.162}$$

ou

$$x_{1-\beta} = x_\alpha - \sqrt{2n}\, \sigma_s^2/\sigma^2. \tag{2.162'}$$

Remarquons que lorsque l'on utilise au lieu du détecteur de la figure 2.7 le corrélateur (fig. 2.8) réalisant l'algorithme

$$\sum_{i=1}^{n} x_i y_i \geqslant C, \tag{2.163}$$

la probabilité $1 - \beta$ de détection correcte d'un signal faible pour $n \gg 1$ et $\alpha$ donné se calcule par la formule *)

$$x_{1-\beta} = x_\alpha - \sqrt{n}\, \sigma_s^2/\sigma^2. \tag{2.164}$$

En comparant avec (2.162') on voit que la grandeur $x_{1-\beta}$ est dans ce cas de quelque 40 % supérieure que pour l'algorithme optimal (2.158).

Lorsque l'on ne connaît pas la valeur moyenne ni la variance du signal et du bruit normaux, le problème envisagé de la détection

---

*) Avec $C = x_\alpha \sigma^2 \sqrt{n}$ et $\beta = F\,[(C - n\sigma_s^2)/(\sigma^2 \sqrt{n})]$.

du signal revient à la vérification de l'hypothèse $r = 0$ contre l'alternative $r > 0$ pour des valeurs quelconques des moyennes et des variances du signal et du bruit. La règle optimale non biaisée de la détection impose dans ce cas la comparaison avec un seuil

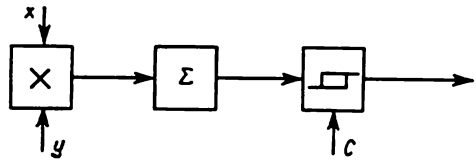



Fig. 2.8. Schéma fonctionnel d'un corrélateur.

de l'estimation du maximum de vraisemblance du coefficient de corrélation (voir [1], § 4.2)

$$\hat{r} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) / [\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2]^{1/2}, \quad (2.165)$$

où

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i.$$

On peut montrer (voir [1], théorème 4.2.6) que la statistique $\hat{r}$ est asymptotiquement normale de moyenne $r$ et de variance $(1 - r^2)^2/n$. Dans ce cas la valeur du seuil $C$ dans l'algorithme

$$\hat{r} \geqslant C \quad (2.166)$$

pour $n \gg 1$ et une probabilité donnée de fausse alarme est

$$C = x_\alpha / \sqrt{n}, \quad (2.167)$$

et on trouve la probabilité de perte du signal pour $r^2 \ll 1$ (signal faible) à partir de la relation des quantiles de la distribution normale :

$$x_{1-\beta} = x_\alpha - r\sqrt{n}. \quad (2.168)$$

Il est évident que pour $\bar{x} = \bar{y} = 0$ et des variances données, (2.165) devient (2.163), et (2.168) s'identifie à (2.164).

**2.5.2. Corrélateur de coïncidence des polarités.** Supposons que les densités de probabilité du signal et du bruit sont données par des fonctions symétriques par rapport à l'origine des coordonnées et que l'on connaît les variances du signal et du bruit et les moments centrés quatre de la distribution du bruit. Désignons respectivement par $w_{11}(x)$, $w_{12}(x)$, $w_{1s}(x)$ les densités de probabilité des

bruits dans le premier et le second canal et la densité de probabilité du signal, par $\sigma_1^2$, $\sigma_2^2$, $\sigma_3^2$ les variances des bruits et du signal et par $M_{41}$, $M_{42}$ les moments centrés quatre des bruits. Le problème de la détection d'un signal stochastique revient à vérifier l'hypothèse

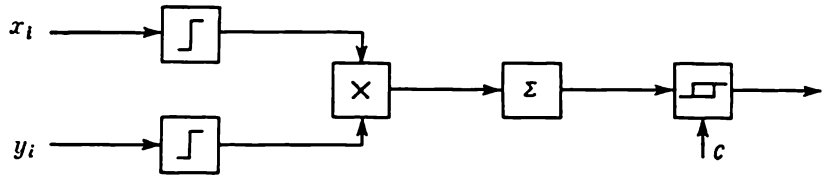



Fig. 2.9. Corrélateur de coïncidence des polarités.

$H$ selon laquelle les processus $x(t)$ et $y(t)$ observés dans le canal sont indépendants, c'est-à-dire que leur distribution mutuelle (aux mêmes instants) est

$$w_2(x, y \mid H) = w_{11}(x)\, w_{12}(y), \tag{2.169}$$

contre l'alternative $K$ que la distribution mutuelle des processus est *)

$$w_2(x, y \mid K) = \int_{-\infty}^{\infty} w_{11}(x-z)\, w_{12}(y-z)\, w_{1s}(z)\, dz. \tag{2.170}$$

Pour la détection d'un signal stochastique noyé dans des bruits additifs indépendants (pour des suppositions mentionnées) nous allons utiliser l'algorithme de signe suivant: la décision que le signal est présent (l'hypothèse $H$ est rejetée) est prise si

$$\sum_{i=1}^{n} \operatorname{sgn} x_i \operatorname{sgn} y_i \geqslant C_1. \tag{2.171}$$

On suppose évidemment que les composantes des vecteurs des observations **x** et **y** sont indépendantes.

L'algorithme (2.171) régit le fonctionnement du *corrélateur de coïncidence des polarités* (appelé parfois corrélateur des polarités) (fig. 2.9).

Compte tenu de la relation existant entre les fonctions sgn $x$ et $u(x)$ (voir (2.39)) on peut écrire l'algorithme (2.171) comme suit:

$$\sum_{i=1}^{n} u(x_i y_i) \geqslant C. \tag{2.172}$$

---

*) La fonction sous l'intégrale (2.170) représente la densité de probabilité mutuelle (tridimensionnelle) des bruits additifs indépendants et du signal.

La somme dans (2.172), égale au nombre de coïncidences des signes des observations dans les canaux, suit la loi binomiale de paramètres $(n, {}^1/_2)$ pour l'hypothèse $H$ et de paramètres $(n, p)$ pour l'alternative $K$, de plus *)

$$p = P\{x > 0,\ y > 0\} + P\{x < 0,\ y < 0\} =$$

$$= \int_{-\infty}^{\infty} F_{11}(z) F_{12}(z) w_{1s}(z)\, dz + \int_{-\infty}^{\infty} [1 - F_{11}(z)]\,[1 - F_{12}(z)]\, w_{1s}(z)\, dz =$$

$$= \frac{1}{2} + 2 \int_{-\infty}^{\infty} \left[F_{11}(z) - \frac{1}{2}\right]\left[F_{12}(z) - \frac{1}{2}\right] w_{1s}(z)\, dz. \qquad (2.173)$$

Avec l'algorithme (2.172) la probabilité de fausse alarme est

$$\alpha = P\left\{\sum_{i=1}^{n} u(x_i y_i) \geqslant C \mid H\right\} = \sum_{k=[C]+1}^{n} \binom{n}{k} (1/2)^k. \qquad (2.174)$$

La formule (2.174), qui est au fond analogue à (2.48), fournit une valeur du seuil $C$ (dans (2.172)) indépendante de la distribution pour des distributions symétriques quelconques des bruits pour une probabilité donnée de fausse alarme. Autrement dit, le seuil établi dans le corrélateur polaire coïncide avec le seuil réglé dans le détecteur de signe.

Ainsi, dans les hypothèses que nous avons faites, le corrélateur de coïncidence des probabilités est un détecteur non paramétrique d'un signal stochastique noyé dans des bruits additifs indépendants.

Pour un échantillon de taille importante la distribution de la somme dans (2.172) tend vers une loi normale de moyenne $np$ et de variance $np(1 - p)$, avec $p = {}^1/_2$ pour l'hypothèse $H$ et $p > {}^1/_2$ pour l'alternative $K$, cette distribution étant donnée par (2.173). Les probabilités de fausse alarme et de perte du signal sont dans ce cas données par les formules (2.50) et (2.51), et la relation entre les quantiles $x_\alpha$ et $x_{1-\beta}$ par la formule (2.52) où l'on porte $p$ calculé d'après (2.173).

**2.5.3. Efficacité relative du corrélateur de coïncidences des polarités.** Comparons le corrélateur de coïncidences des polarités avec le détecteur d'un signal stochastique, optimal pour des distributions normales du signal et du bruit (voir (2.158)). Supposons que l'algorithme (2.158) est utilisé pour des distributions symétriques quelconques du signal et des bruits stationnaires additifs indépen-

*) Les paramètres de la loi binomiale sont: le nombre $n$ d'expériences indépendantes dans le schéma de Bernoulli et la probabilité $p$ de réalisaton de l'événement dans chaque expérience.

dants. Dans ce cas pour l'hypothèse $H$ (il n'y a pas de signal) on a

$$m_1\{\sum_{i=1}^{n}(x_i+y_i)^2 \mid H\} = n(\sigma_1^2+\sigma_2^2), \tag{2.175}$$

$$M_2\{\sum_{i=1}^{n}(x_i+y_i)^2 \mid H\} = n[M_{41}+M_{42}+6\sigma_1^2\sigma_2^2-(\sigma_1^2+\sigma_2^2)^2] =$$
$$= n(M_{41}+M_{42}+4\sigma_1^2\sigma_2^2-\sigma_1^4-\sigma_2^4). \tag{2.176}$$

Pour l'alternative $K$ (le signal est présent) on a

$$m_1\{\sum_{i=1}^{n}(x_i+y_i)^2 \mid K\} = n(\sigma_1^2+\sigma_2^2+4\sigma_s^2), \tag{2.177}$$

et la variance pour un signal faible ($\sigma_s^2 \ll \sigma_1^2+\sigma_2^2$) est donnée par (2.176). Les grandeurs $M_{41}$, $M_{42}$, $\sigma_1^2$, $\sigma_2^2$, $\sigma_s^2$ sont respectivement les moments quatre, les variances des bruits et la variance du signal.

La statistique dans (2.158) est asymptotiquement normale, ses paramètres étant donnés par (2.175) à (2.177). Le seuil est maintenant déterminé pour $n \gg 1$ et une probabilité donnée de fausse alarme par l'égalité

$$C = x_\alpha\sqrt{n}(M_{41}+M_{42}+4\sigma_1^2\sigma_2^2-\sigma_1^4-\sigma_2^4)^{1/2}+n(\sigma_1^2+\sigma_2^2), \tag{2.178}$$

et la probabilité de perte du signal se trouve à partir de la relation

$$x_{1-\beta} = x_\alpha - 4\sigma_s^2\sqrt{2}(M_{41}+M_{42}+4\sigma_1^2\sigma_2^2-\sigma_1^4-\sigma_2^4)^{-1/2}. \tag{2.179}$$

Pour des bruits soumis à la distribution normale on a $M_{41} = M_{42} = M = 3\sigma^4$ pour $\sigma_1^2 = \sigma_2^2 = \sigma^2$ et les formules (2.178) et (2.179) coïncident avec (2.160) et (2.162) respectivement.

Pour un signal faible, on obtient à partir de (2.173), compte tenu de $F_{1i}(z) = {}^1/_2 + z\,w_{1i}(z) + \ldots$, $i = 1, 2$,

$$2p-1 = 4\int_{-\infty}^{\infty}[zw_{11}(0)+\ldots][zw_{12}(0)+\ldots]\,w_{1s}(z)\,dz =$$
$$= 4\sigma_s^2 w_{11}(0)\,w_{12}(0)+o(\sigma_s^2), \tag{2.180}$$
$$4p(1-p) = 1-o(\sigma_s^2). \tag{2.180'}$$

En substituant (2.180) et (2.180′) dans (2.67) on obtient dans le cas d'utilisation de l'algorithme (2.172) pour $n \to \infty$ et $\sigma_s^2\sqrt{n} = \text{const}$:

$$x_{1-\beta} = x_\alpha - 4\sigma_s^2 w_{11}(0)\,w_{12}(0)\sqrt{n}. \tag{2.181}$$

(2.179) et (2.181) fournissent immédiatement l'expression du coefficient d'efficacité relative du corrélateur de coïncidence des polarités par rapport au détecteur optimal pour des bruits normale-

ment répartis, soit

$$\rho = w_{11}^2(0)\, w_{12}^2(0)\,(M_{41} + M_{42} + 4\sigma_1^2\sigma_2^2 - \sigma_1^4 - \sigma_2^4). \qquad (2.182)$$

Si les distributions des bruits sont normales et de même variance, on a

$$w_{11}^2(0) = w_{12}^2(0) = 1/(2\pi\sigma^2), \qquad M_{41} = M_{42} = 3\sigma^4$$

et en vertu de (2.182)

$$\rho = 8\sigma^4/(2\pi\sigma^2)^2 = 2/\pi^2 \approx 0{,}2, \qquad (2.183)$$

c'est-à-dire le corrélateur de coïncidence des polarités est bien moins efficace que le détecteur optimal d'un signal stochastique.

Cependant pour des bruits laplaciens (voir (2.72)) on a

$$w_{11}^2(0) = w_{12}^2(0) = \lambda^2/4 = 1/(2\sigma^2), \qquad M_{41} = M_{42} = 6\sigma^4$$

et à partir de (2.182) on trouve

$$\rho = 14\sigma^4/(2\sigma^2)^2 = 3{,}5. \qquad (2.184)$$

Pour des bruits uniformément distribués sur l'intervalle $|x| \leqslant \sigma\sqrt{3}$ on a

$$w_{11}^2(0) = w_{12}^2(0) = 1/(12\sigma^2), \qquad M_{41} = M_{42} = 19\sigma^4/5$$

et

$$\rho = \frac{19\sigma^4}{5\,(12\sigma^2)^2} \approx 0{,}03, \qquad (2.185)$$

c'est-à-dire que dans ce cas l'efficacité du corrélateur de coïncidence des polarités est bien plus petite que dans le cas de l'algorithme optimal.

Enfin pour un bruit sinusoïdal de phase aléatoire (voir (2.74)) on a

$$w_{11}^2(0) = w_{12}^2(0) = 1/(2\pi^2\sigma^2), \qquad M_{41} = M_{42} = 3\sigma^4/2$$

et

$$\rho = 5\sigma^4/(2\pi^2\sigma^2)^2 = 5/4\pi^4 \approx 0{,}01. \qquad (2.186)$$

Si $w_1(0) = 0$, c'est-à-dire si la densité de probabilité des bruits est nulle à l'origine des coordonnées (distributions symétriques multimodales), on a en vertu de (2.182) $\rho = 0$.

En utilisant (2.164) et (2.181) on trouve le coefficient d'efficacité asymptotique relative du corrélateur de coïncidence des polarités par rapport au corrélateur ordinaire:

$$\rho = 16\sigma^4 w_{11}^2(0)\, w_{12}^2(0). \qquad (2.187)$$

La relation (2.187) donne pour un bruit à distribution normale $\rho \approx 0{,}4$ et pour un bruit à distribution laplacienne $\rho = 4$.

**2.5.4. Corrélateur modifié de coïncidence des polarités.** Supposons qu'on connaît les quantiles de la distribution du bruit $z_{q_j} = F_1^{-1}(q_j)$, $j = 1, \ldots, m$. On peut utiliser cette information supplémentaire pour améliorer la qualité de l'algorithme (2.171), tout comme on l'avait fait dans 2.3.7 pour accroître l'efficacité du détecteur de signe. L'algorithme modifié sera

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \operatorname{sgn}(x_i/\sigma - z_{q_j}) \operatorname{sgn}(y_i/\sigma - z_{q_j}) \geqslant C, \tag{2.188}$$

où $\sigma^2$ est la variance du bruit.

Nous appellerons *corrélateur modifié de coïncidence des polarités* le dispositif réalisant l'algorithme (2.188). Supposant que les bruits indépendants suivent une même loi de répartition, on peut montrer que le coefficient d'efficacité asymptotique relative de l'algorithme (2.188) par rapport à l'algorithme (2.158), optimal pour des bruits additifs normalement répartis, est égal à [31]

$$\rho = 2(M_4 + \sigma^4) \left[\sum_{j=1}^{m} w_1^2(\sigma z_{q_j})\right]^2 \Big/ \sum_{\substack{i,j=1 \\ i<j}}^{m} \sigma_{ij}^2, \tag{2.189}$$

avec

$$\sigma_{ij}^2 = \{(1-2q_i) + (1-2q_j) - 1\}^2 - \{(2q_j - 1)(2q_i - 1)\}^2, \tag{2.189'}$$

$M_4$ étant le moment centré quatre de la distribution du bruit.

Dans la table 2.4 sont indiquées les valeurs du coefficient d'efficacité asymptotique relative obtenues dans [31] pour le corrélateur modifié de coïncidence des polarités par rapport au détecteur optimal de Neyman-Pearson, ainsi que par rapport au corrélateur de coïncidence des polarités (2.172) et au corrélateur ordinaire (2.163).

*Table 2.4*

| Distribution | Détecteur de Neyman-Pearson | Corrélateur de coïncidence des polarités | Corrélateur ordinaire |
|---|---|---|---|
| Normale | 0,285 | 1,41 | 0,57 |
| de Laplace | 3,43 | 0,98 | 6,85 |

**2.5.5. Algorithmes de rang de détection d'un signal stochastique.** Soient $\mathbf{x} = (x_1, \ldots, x_n)$ et $\mathbf{y} = (y_1, \ldots, y_n)$ des échantillons des observations dans deux canaux, et soient $\mathbf{R} = (R_1, \ldots, R_n)$ et $\mathbf{Q} = (Q_1, \ldots, Q_n)$ les vecteurs de rang de ces échantillons. Pour vérifier l'hypothèse $H$ : il n'y a pas de signal (les échantillons $\mathbf{x}$ et $\mathbf{y}$ sont indépendants) contre l'alternative $K$ : le signal est présent dans les deux canaux (les échantillons $\mathbf{x}$ et $\mathbf{y}$ sont liés) on peut uti-

liser la règle suivante: on adopte l'alternative $K$ si

$$S_n(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} R_i Q_i \geqslant C, \tag{2.190}$$

et on la rejette si c'est l'inégalité inverse de (2.190) qui est vérifiée.

La statistique $S_n$ dans (2.190) est appelée *coefficient de corrélation de rang de Spearmen*. La valeur moyenne et la variance de cette statistique pour l'hypothèse $H$ sont

$$m_1\{S_n(\mathbf{x}, \mathbf{y}) \mid H\} = n(n+1)^2/4,$$
$$M_2\{S_n(\mathbf{x}, \mathbf{y})]H\} = n^2(n+1)^2(n-1)/144. \tag{2.191}$$

La statistique $S_n$ étant asymptotiquement normale, pour $n \gg 1$ le seuil dans (2.190) est donné par

$$C = (x_\alpha + 3\sqrt{n})\, n^{5/2}/12,$$

où $\alpha$ est la probabilité donnée de fausse alarme et $x_\alpha$ l'$\alpha$-quantile de la loi normale.

L'algorithme utilisant la *statistique de Kendall* est équivalent par son efficacité à l'algorithme (2.190) (voir la table dans A.3.3 empruntée à [20])

$$S_n(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \operatorname{sgn}(x_i - x_j) \operatorname{sgn}(y_i - y_j). \tag{2.192}$$

Le schéma du détecteur de signaux stochastiques basé sur le calcul de la statistique (2.190) est donné sur la figure 2.10.

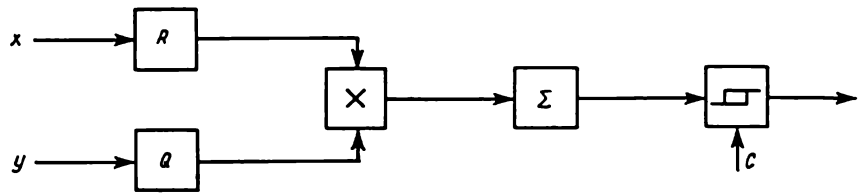



Fig. 2.10. Schéma fonctionnel d'un détecteur de rang d'un signal stochastique.

On peut montrer que pour des bruits normalement répartis les coefficients d'efficacité asymptotique relative des algorithmes de rang de détection de signaux stochastiques sont égaux à $9/(2\pi^2) \approx \approx 0{,}45$ par rapport à l'algorithme (2.158), optimal pour des bruits normalement répartis, à 2,25 par rapport au corrélateur de coïncidence des polarités, à $9/\pi^2 \approx 0{,}91$ par rapport au corrélateur ordinaire et à 1,6 par rapport au corrélateur modifié de coïncidence des polarités (pour trois quantiles donnés de la répartition des bruits).

Pour calculer la statistique (2.192) il faudra effectuer environ $n^2/2$ opérations de comparaison des échantillons observés. On peut diminuer notablement le nombre des opérations sans perdre sensiblement en efficacité de l'algorithme de détection, en utilisant la statistique mixte (voir 2.4.3)

$$S_n(\mathbf{x}, \mathbf{y}) = \sum_{i=0}^{r-1} S_n^{(i)}(\mathbf{x}, \mathbf{y}), \tag{2.193}$$

où

$$S_n^{(i)}(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^{m} \sum_{k=j+1}^{m} \operatorname{sgn}(x_{im+j} - x_{im+k}) \operatorname{sgn}(y_{im+j} - y_{im+k}). \tag{2.194}$$

Pour des bruits normalement répartis le coefficient d'efficacité asymptotique relative du détecteur utilisant une statistique mixte de la forme (2.193) par rapport au corrélateur de coïncidence des polarités est

$$\rho = \frac{9}{2} \frac{m-1}{2m+5}. \tag{2.195}$$

Pour $m \to \infty$, on a $\rho \to 9/4$, ce qui correspond au cas de la statistique (2.192). Si l'on répartit les échantillons en 10 groupes, le nombre d'opérations pour $n = 100$ diminue de 10 fois, et le coefficient d'efficacité asymptotique relative $\rho = 81/50 = 1{,}62$, soit 72% du coefficient correspondant à la statistique (2.192).

## 2.6. ALGORITHME NON PARAMÉTRIQUE DE DÉTECTION D'UN SIGNAL EN PRÉSENCE D'UN ÉCHANTILLON D'APPRENTISSAGE DU BRUIT

**2.6.1. Position du problème.** Revenons au problème de la détection du signal constant $s(t) \equiv a$ noyé dans un bruit additif. Dans 2.3.1 nous avons supposé que la densité de probabilité $w_1(x \mid H) = w_1(x)$ du bruit était symétrique par rapport à l'origine des coordonnées. Ici nous n'avons pas besoin de cette hypothèse. Supposons cependant que l'observateur dispose à l'avance d'un échantillon indépendant $\mathbf{y} = (y_1, \ldots, y_{n_1})$ du bruit. Nous dirons que c'est un *échantillon d'apprentissage*. Par rapport à un autre échantillon indépendant $\mathbf{x} = (x_1, \ldots, x_n)$ il y a lieu de vérifier l'hypothèse $H$ selon laquelle c'est également un échantillon de distribution $w_1(x)$ contre l'alternative $K$ que c'est un échantillon du mélange du signal et du bruit de distribution $w_1(x - a)$. Ainsi formulé, le problème coïncide avec celui de 2.1.7 où il s'agissait de vérifier l'hypothèse relativement à l'appartenance de deux échantillons (d'apprentissage et observé) à une même distribution.

**2.6.2. Algorithme optimal de détection dans le cas d'un bruit normal additif.** Nous allons tout d'abord trouver l'algorithme optimal de détection d'un signal constant noyé dans un bruit normal additif de moyenne inconnue $a_0$ et de puissance moyenne $\sigma^2$ donnée. Autrement dit, il faut trouver la procédure optimale de vérification de l'hypothèse selon laquelle deux échantillons indépendants y et x appartiennent à une distribution normale de même moyenne $a = a_0$, contre l'alternative $K$ que $a \neq a_0$, sous la condition que la variance $\sigma^2$ de la distribution normale est donnée *). Supposons que $a > a_0$. En utilisant au lieu de la valeur moyenne inconnue $a_0$ l'estimation du maximum de vraisemblance $\hat{a}_0 = \frac{1}{n_1}\sum_{i=1}^{n_1} y_i$, calculée d'après l'échantillon d'apprentissage $y_1, \ldots, y_{n_1}$, on obtient l'algorithme suivant, optimal d'après le critère de Neyman-Pearson : pour une probabilité donnée de fausse alarme $\alpha$, on décide que le signal est présent (l'hypothèse $H$ est rejetée) si

$$\bar{x} - \bar{y} \geqslant x_\alpha \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \tag{2.196}$$

où $\bar{x} = \frac{1}{n_2}\sum_{i=1}^{n_2} x_i$, $\bar{y} \equiv \hat{a}_0$ et $x_\alpha$ est l'$\alpha$-quantile de la loi normale (voir tome II, (1.68) et (1.76)).

Si la variance du bruit $\sigma^2$ est inconnue, en utilisant l'estimation du maximum de vraisemblance établie d'après l'échantillon composé (y, x) on arrive à l'algorithme suivant (voir tome II, (1.157)) :

$$t(\mathrm{x}, \mathrm{y}) = \frac{(\bar{x} - \bar{y}) \Big/ \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}}{\left\{\frac{1}{n_1 + n_2 - 1}\left[\sum_{i=1}^{n_2} (x_i - \bar{x})^2 + \sum_{i=1}^{n_1} (y_i - \bar{y})^2\right]\right\}^{1/2}} \geqslant C. \tag{2.197}$$

Pour l'alternative $K$ la statistique $t$ (x, y) suit une loi de Student non centrée à $(n_1 + n_2 - 2)$ degrés de liberté et d'écart

$$\delta = (a - a_0)/\sigma(1/n_1 + 1/n_2)^{1/2}. \tag{2.197'}$$

Pour l'hypothèse $H$ on a $\delta = 0$, ce qui correspond à la distribution de Student centrée, donc le seuil $C$ dans (2.197) pour une probabilité donnée $\alpha$ de fausse alarme coïncide avec l'$\alpha$-quantile de cette distribution, c'est-à-dire $C = t_\alpha$.

---

*) C'est un cas particulier du problème de Berens-Fisher qui consiste à vérifier l'égalité des moyennes de deux distributions normales de variances inconnues (voir [12]).

Remarquons que la statistique $t$ (x, y) est asymptotiquement équivalente à la statistique $(\bar{x}-\bar{y})\Big/\left(\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$, qui pour une *distribution continue quelconque* du bruit est asymptotiquement normale de moyenne égale à $\delta$ et de variance unité.

Si l'alternative $K$ est bilatérale, i.e. $a \neq a_0$ sans avoir $a > a_0$, les égalités (2.196) et (2.197) deviennent des inégalités (voir tome II, (1.131) et (1.56)), soit :

$$|\bar{x}-\bar{y}| \geqslant x_{\alpha/2}\sigma\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{1/2}, \tag{2.198}$$

$$|t(\mathrm{x}, \mathrm{y})| \geqslant t_{\alpha/2}. \tag{2.199}$$

**2.6.3. Algorithme de Wilcoxon à deux échantillons.** Considérons maintenant un algorithme non paramétrique de détection d'un signal constant noyé dans un bruit additif avec une densité continue quelconque $w_1(x)$, faisant usage d'un échantillon d'apprentissage du bruit. Cet algorithme utilise la statistique U (x, y), égale au nombre total des cas où l'élément de l'échantillon indépendant observé dépasse l'élément correspondant de l'échantillon d'apprentissage indépendant. En vertu de la terminologie adoptée dans 2.1.7, la statistique U (x, y) est égale à la somme des inversions formées par les éléments de l'échantillon observé avec ceux de l'échantillon d'apprentissage, c'est-à-dire

$$\mathrm{U}(\mathrm{x}, \mathrm{y}) = \sum_{i=1}^{n_2}\sum_{j=1}^{n_1} u(x_i - y_j). \tag{2.200}$$

La valeur moyenne et la variance de la statistique U (x, y) pour l'hypothèse $H$ sont (voir (2.36'))

$$m_1\{\mathrm{U}(\mathrm{x}, \mathrm{y}) \mid H\} = n_1 n_2/2,$$

$$M_2\{\mathrm{U}(\mathrm{x}, \mathrm{y}) \mid H\} = (1/12)\,[n_1 n_2 (n_1 + n_2 + 1)]. \tag{2.200'}$$

La distribution de la statistique U (x, y) pour l'hypothèse $H$ est symétrique par rapport à la valeur moyenne et *ne dépend pas de la distribution du bruit* $w_1(x)$. Supposons que la valeur moyenne du bruit est nulle. L'algorithme unilatéral en vertu duquel on prend la décision que le signal $a > 0$ est présent si

$$\mathrm{U}(\mathrm{x}, \mathrm{y}) \geqslant C_1 \tag{2.201}$$

sera non paramétrique car le seuil $C_1$ ne dépend pas de la distribution du bruit et est déterminé par une probabilité donnée $\alpha$ de fausse alarme

$$P\{\mathrm{U}(\mathrm{x}, \mathrm{y}) \geqslant C_1 \mid H\} = \alpha. \tag{2.202}$$

La figure 2.11 schématise un dispositif de détection réalisant l'algorithme (2.201) appelé algorithme unilatéral de Wilcoxon à deux échantillons.

Comme nous l'avons noté dans 2.1.7, la statistique U (x, y) est asymptotiquement normale pour $n_1 \to \infty$, $n_2 \to \infty$, une bonne

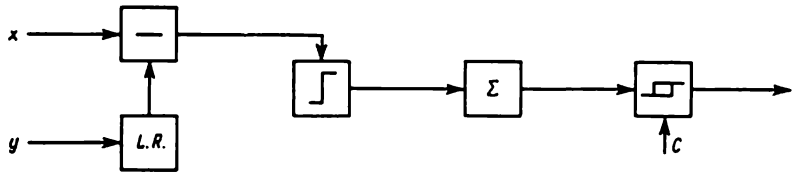



Fig. 2.11. Schéma fonctionnel d'un détecteur de signal utilisant l'algorithme de Wilcoxon à deux échantillons.

approximation de la loi normale étant observée déjà pour $n_1 > 3$, $n_2 > 20$. Pour ces conditions de normalisation, on obtient à partir de (2.202) la grandeur du seuil, à savoir

$$C_1 = x_\alpha [n_1 n_2 (n_1 + n_2 + 1)/12]^{1/2} + n_1 n_2/2, \tag{2.203}$$

$x_\alpha$ étant l'$\alpha$-quantile de la loi normale.

Compte tenu de (2.41), on peut exprimer la statistique U (x, y) en fonction des rangs $R_i$ des éléments $x_i$ de l'échantillon observé dans l'échantillon *composé* (x, y) :

$$\mathrm{U}(\mathrm{x}, \mathrm{y}) = \sum_{i=1}^{n_2} (R_i - i) = \sum_{i=1}^{n_2} R_i - \frac{n_2(n_2+1)}{2}. \tag{2.204}$$

Il est également évident que la statistique (2.200) ne diffère que par une constante de la statistique

$$\mathrm{U}_1(\mathrm{x}, \mathrm{y}) = \sum_{i=1}^{n_2} \sum_{j=1}^{n_1} \operatorname{sgn}(x_i - y_j). \tag{2.205}$$

Lors de la détection d'un signal constant de signe quelconque on utilise l'algorithme bilatéral non paramétrique de Wilcoxon à deux échantillons, en vertu duquel on prend la décision sur la présence du signal si

$$\left| \mathrm{U}(\mathrm{x}, \mathrm{y}) - \frac{n_1 n_2}{12} \right| \geqslant C_2, \tag{2.206}$$

le seuil $C_2$ pour des échantillons de tailles importantes étant égal à

$$C_2 = x_{\alpha/2} [n_1 n_2 (n_1 + n_2 + 1)/12]^{1/2}. \tag{2.207}$$

**2.6.4. Efficacité asymptotique relative de l'algorithme de Wilcoxon à deux échantillons.** Comparons l'algorithme (unilatéral) de Wilcoxon à deux échantillons avec l'algorithme de détection de signaux optimal en présence de bruits normaux. Supposons que l'algorithme de détection (2.196) ou l'algorithme qui lui est asymptotiquement équivalent (2.197) sont utilisés pour des bruits additifs de distribution quelconque. Cherchons le coefficient d'efficacité asymptotique relative de l'algorithme (2.201) par rapport à l'algorithme (2.196).

Comme

$$m_1\{\bar{x}-\bar{y}\,|\,K\}=a-a_0,\quad M_2\{\bar{x}-\bar{y}\,|\,H\}=\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right),$$

on a

$$\left[\frac{d}{da}\,m_1\{\bar{x}-\bar{y}\,|\,K\}\right]^2\Big/M_2\{\bar{x}-\bar{y}\,|\,H\}=\frac{n_1n_2}{(n_1+n_2)\,\sigma^2}. \qquad (2.208)$$

Calculons la valeur moyenne de la statistique U (x, y) pour l'alternative $K$. Avec (2.200) on trouve

$$m_1\{\mathrm{U}(\mathrm{x},\ \mathrm{y})\,|\,K\}=\sum_{i=1}^{n_2}\sum_{j=1}^{n_1} m_1\{u(x_i-y_j)\,|\,K\}=n_1n_2p, \qquad (2.209)$$

où $p$ est la probabilité que l'élément de l'échantillon soit supérieur à l'élément correspondant de l'échantillon d'apprentissage, i.e.

$$p=P\{x>y\,|\,K\}=\int_{-\infty}^{\infty} F_1(z-a)\,w_1(z)\,dz, \qquad (2.210)$$

avec

$$F_1(z)=\int_{-\infty}^{z} w_1(u)\,du.$$

On tire de (2.200′), (2.209) et (2.210)

$$\left[\frac{d}{da}\,m_1\{\mathrm{U}(\mathrm{x},\ \mathrm{y})\,|\,K\}_{a=0}\right]^2\Big/M_2\{\mathrm{U}(\mathrm{x},\ \mathrm{y})\,|\,H\}=$$

$$=\frac{12n_1^2n_2^2}{n_1n_2(n_1+n_2+1)}\left(\int_{-\infty}^{\infty} w_1^2(z)\,dz\right)^2. \qquad (2.211)$$

L'expression du coefficient d'efficacité asymptotique relative cherché se déduit directement à partir de (2.208) et (2.211), soit

$$\rho=12\sigma^2\left(\int_{-\infty}^{\infty} w_1^2(z)\,dz\right)^2. \qquad (2.212)$$

Notons que (2.212) coïncide avec l'expression (2.102) du coefficient d'efficacité asymptotique relative de l'algorithme signe-rang de détection d'un signal constant noyé dans un bruit additif à distribution symétrique, par rapport à l'algorithme linéaire. Donc tous les cas particuliers évoqués dans 2.4.2 et toutes les valeurs numériques du coefficient $\rho$ qui y sont données se rapportent automatiquement à l'algorithme de Wilcoxon à deux échantillons envisagé ici.

## 2.7. ALGORITHMES DE SIGNE ET DE RANG DE DÉTECTION DES SIGNAUX NOYÉS DANS DES BRUITS CORRÉLÉS

**2.7.1. Modèle autorégressif des bruits corrélés.** Supposons que l'échantillon du bruit $\mathbf{x} = (x_1, \ldots, x_n)$ représente une suite autorégressive aléatoire stationnaire du premier ordre, c'est-à-dire (voir par exemple [32])

$$x_1 = \varepsilon_1, \quad x_j = \lambda x_{j-1} + \varepsilon_j, \quad j = 2, \ldots, n, \tag{2.213}$$

où $\varepsilon_1, \ldots, \varepsilon_n$ est une suite de grandeurs aléatoires indépendantes de même distribution de moyenne nulle et de variance finie, avec $|\lambda| < 1$. En vertu de (2.213) on a

$$x_m = \sum_{l=0}^{m-1} \lambda^l \varepsilon_{m-l} = \sum_{l=1}^{m} \lambda^{m-l} \varepsilon_l, \quad m = 1, \ldots, n, \tag{2.214}$$

et

$$x_{m+k} = \sum_{l=0}^{m+k-1} \lambda^l \varepsilon_{m+k-l} = \sum_{l=0}^{k-1} \lambda^l \varepsilon_{m+k-l} + \sum_{l=k}^{m+k-1} \lambda^l \varepsilon_{m+k-l} =$$
$$= \sum_{l=0}^{k-1} \lambda^l \varepsilon_{m+k-l} + \sum_{l=0}^{m-1} \lambda^{l+k} \varepsilon_{m-l}, \quad k \geqslant 1,$$

ou

$$x_{m+k} = \lambda^k x_m + \eta_{m+k}, \quad k \geqslant 1, \quad m = 1, \ldots, n, \tag{2.215}$$

avec

$$\eta_{m+k} = \sum_{l=0}^{k-1} \lambda^l \varepsilon_{m+k-l}. \tag{2.215'}$$

Il est clair que pour $k \geqslant 1$ les grandeurs $x_m$ et $\eta_{m+k}$ sont indépendantes. Dans le cas particulier ($k = 1$) les grandeurs $x_m$ et $\varepsilon_{m+1}$, $m = 1, \ldots, n-1$, sont indépendantes. Il est facile maintenant d'indiquer la condition à laquelle doit satisfaire la distribution des variables $\varepsilon_i$ pour que tous les éléments de l'échantillon du bruit $(x_1, \ldots, x_n)$ appartiennent à une même distribution $w_1(x)$. Soient $\theta_1(v)$ la fonction caractéristique correspondant à la densité de probabilité $w_1(x)$, et $\Theta_{\eta_{m+n}}(v)$ la fonction caractéristique

de la variable aléatoire $\eta_{m+k}$. Les variables $x_m$ et $\eta_{m+k}$ étant indépendantes, on obtient la condition cherchée à partir de (2.215)

$$\theta_1(v) = \theta_1(\lambda^{2k}v)\,\Theta_{\eta_{m+k}}(v)$$

ou

$$\Theta_{\eta_{m+k}}(v) = \theta_1(v)/\theta_1(\lambda^{2k}v). \tag{2.216}$$

La fonction de corrélation de la suite (2.213) (voir (2.215)) est

$$B_x(k) = m_1\{x_m x_{m+k}\} = m_1\{x_m(\lambda^k x_m + \eta_{m+k})\} = \\ = \lambda^k m_1\{x_m^2\} + m_1\{x_m \eta_{m+k}\}.$$

Comme $m_1\{x_m\eta_{m+k}\} = 0$, en introduisant la désignation $\sigma^2 = m_1\{x_i^2\}$ on obtient

$$B_x(k) = B_x(-k) = \sigma^2\lambda^k. \tag{2.217}$$

En vertu de (2.215) la suite $\{x_m\}$ est un processus markovien ergodique simplement connexe à temps discret (voir tome I, § 4.5).

**2.7.2. Algorithme de signe compte tenu de l'autorégression.** Considérons de nouveau l'algorithme de signe (2.47) de détection d'un signal constant noyé dans un bruit à distribution symétrique, mais les valeurs échantillonnées du bruit ne seront pas supposées indépendantes comme dans 2.3.1, elles seront conformes au modèle d'autorégression introduit plus haut.

La corrélation entre les valeurs échantillonnées du bruit *ôte à l'algorithme de signe* (2.47) *la propriété d'être non paramétrique.* Pour nous en rendre compte nous allons montrer que dans le cas envisagé le seuil $C$ dans (2.47) *dépend* de la distribution du bruit. Déterminons la valeur moyenne et la variance de la statistique de signe $\sum_{i=1}^{n} u(x_i)$ pour l'hypothèse $H$. La moyenne de la somme de variables aléatoires étant toujours égale à la somme des moyennes, on a

$$m_1\left\{\sum_{i=1}^{n} u(x_i)|H\right\} = \sum_{i=1}^{n} m_1\{u(x_i)|H\} = \sum_{i=1}^{n} P\{x_i > 0|H\} = n/2, \tag{2.218}$$

où $P\{x_i > 0 \mid H\} = {}^1/_2$ grâce à la symétrie de la distribution du bruit. La variance de la statistique de signe est

$$M_2\left\{\sum_{i=1}^{n} u(x_i)|H\right\} = m_1\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}\left[u(x_i) - \frac{1}{2}\right]\left[u(x_j) - \frac{1}{2}\right]|H\right\} = \\ = \sum_{i=1}^{n}\sum_{j=1}^{n}[m_1\{u(x_i)\,u(x_j)|H\} - {}^1/_4]. \tag{2.219}$$

Le bruit étant stationnaire, on a

$$B_u(j-i) = m_1\{u(x_i)\,u(x_j)\,|\,H\} - {}^1/_4 = B_u(i-j). \qquad (2.220)$$

En substituant (2.220) dans (2.219) on trouve

$$M_2\Big\{\sum_{i=1}^{n} u(x_i)\,|H\Big\} = \sum_{k=-n}^{n} (n-|k|)\,B_u(k) =$$

$$= \frac{n}{4} + 2\sum_{k=1}^{n-1} (n-k)\,B_u(k). \qquad (2.221)$$

Pour le modèle autorégressif (2.213), la fonction de corrélation $B_u(k)$ peut être exprimée en fonction de la densité de probabilité $w_1(x)$ de la distribution du bruit. En vertu de (2.220) et (2.215) on a

$$B_u(k) + \frac{1}{4} = P\{x_m > 0,\ x_{m+k} > 0\,|\,H\} =$$

$$= P\{x_m > 0,\ \lambda^k x_m + \eta_{m+k} > 0\,|\,H\} =$$

$$= \int_0^\infty w_1(x) \int_{-\lambda^k x}^{\infty} w_{\eta_{m+k}}(z)\,dz\,dx, \qquad (2.222)$$

$w_{\eta_{m+k}}(z)$ étant donné par la transformation inverse de Fourier à partir de (2.216), où $\theta_1(v)$ est la fonction caractéristique du bruit.

Ainsi la variance de la statistique de signe des observations pour un bruit corrélé dépend de la distribution du bruit. C'est pourquoi le seuil dans l'algorithme de signe (2.47) dépend de la distribution du bruit et par conséquent cet algorithme cesse d'être non paramétrique.

**2.7.3. Normalité asymptotique de la statistique.** Pour un bruit corrélé du type autorégressif (2.213) la distribution de la statistique de signe $\sum_{i=1}^{n} u(x_i)$ est asymptotiquement normale (pour $n \to \infty$). Cette propriété intéressante de la somme de variables aléatoires liées découle du théorème de la limite centrale pour des suites stationnaires à *mélange intense* (voir [8], théorème 18.5.1). Les suites markoviennes ergodiques sont des suites à mélange intense. La suite $\{x_i\}$ des valeurs échantillonnées étant une suite markovienne ergodique, et $u(x)$ une fonction monotone, la suite $u(x_i)$ est markovienne et ergogique. On peut appliquer le théorème de la limite centrale à la suite $\{y_i\}$ si

$$\lim_{n\to\infty} M_2\Big\{\sum_{i=1}^{n} y_i\Big\} = \infty,$$

$$M_2\{|y_i|^{2+\delta}\} < \infty \text{ pour un certain } \delta > 0. \tag{2.223}$$

En vertu de (2.221) la première condition (2.223) est toujours vérifiée. Pour l'hypothèse $H$ et $\delta = 2$ on a

$$M_2\{[u(x_i)]^4\} = m_1\{([u(x_i) - {}^1/_2]^4\} = ({}^1/_2)^4 P\{x_i \geqslant 0\} + \\ + (-{}^1/_2)^4 P\{x_i < 0\} = {}^1/_{16},$$

donc la seconde condition (2.223) est également vérifiée.

La statistique de signe étant asymptotiquement normale pour un bruit corrélé du type autorégressif, on trouve pour $n \gg 1$ la grandeur du seuil $C$ dans l'algorithme (2.47) pour une probabilité donnée $\alpha$ de fausse alarme:

$$C = x_\alpha [M_2\{\sum_{i=1}^{n} u(x_i)|H\}]^{1/2}. \tag{2.224}$$

**2.7.4. Efficacité relative de l'algorithme de signe de détection d'un signal pour un bruit corrélé.** La valeur moyenne de la statistique de signe pour un bruit corrélé est égale à la valeur moyenne de cette statistique pour un bruit non corrélé (tant pour l'hypothèse $H$ que pour l'alternative $K$) car la moyenne de la somme de variables aléatoires est toujours égale à la somme des moyennes des composantes. Par conséquent, le coefficient d'efficacité asymptotique relative de l'algorithme de signe de détection d'un signal constant noyé dans un bruit corrélé du type autorégressif, par rapport à l'algorithme de signe de détection d'un signal noyé dans un bruit indépendant, est égal à (voir (1.68′) et (1.70′))

$$\rho = \lim_{n\to\infty} M_2\left\{\sum_{i=1}^{n} u(x_i)|H,\ \lambda = 0\right\} \Big/ M_2\left\{\sum_{i=1}^{n} u(x_i)|H,\ \lambda > 0\right\} = \\ = \lim_{n\to\infty} \frac{n}{4}\left[M_2\left\{\sum_{i=1}^{n} u(x_i)|H,\ \lambda > 0\right\}\right]^{-1}. \tag{2.225}$$

Servons-nous de (2.225) pour trouver le coefficient d'efficacité asymptotique relative dans le cas de bruits à distribution normale et laplacienne.

Dans le premier cas la fonction caractéristique du bruit est

$$\theta_1(v) = e^{-\sigma^2 v^2/2}$$

et en vertu de (2.216)

$$\Theta_{\eta_{m+k}}(v) = \exp\{-\sigma^2(1 - \lambda^{2k})v^2/2\}$$

et

$$w_{\eta_{m+k}}(z) = [2\pi(1 - \lambda^{2k})\sigma^2]^{-1/2} \exp\{-z^2[2\sigma^2(1 - \lambda^{2k})]\}. \tag{2.226}$$

En substituant (2.226) dans (2.222) on trouve la fonction de corrélation

$$B_u(k) = \int_0^\infty w(u)\, F\left(-\frac{\lambda^k u}{\sqrt{1-\lambda^{2k}}}\right) du - {}^1/_4, \qquad (2.227)$$

où $w(u)$, $F(u)$ sont la densité de probabilité et la fonction de répartition de la loi normale. La dérivée par rapport au paramètre $\vartheta = \lambda^k$ de l'intégrale

$$\mathrm{I}(\vartheta) = \int_0^\infty w(u)\, F\left(-\frac{\vartheta u}{\sqrt{1-\vartheta^2}}\right) du = \frac{1}{2} - \int_0^\infty w(u)\, F(\vartheta u/\sqrt{1-\vartheta^2})\, du$$

est égale à

$$\mathrm{I}'(\vartheta) = 1/(2\pi\sqrt{1-\vartheta^2}), \qquad (2.228)$$

avec $\mathrm{I}(0) = 1/4$. A partir de (2.227) et (2.228) on trouve *)

$$B_u(k) = \frac{1}{2\pi} \arcsin \lambda^k. \qquad (2.229)$$

Substituons (2.229) dans (2.221), il vient

$$M_2\left\{\sum_{i=1}^{n} u(x_i) \mid H\right\} = \frac{n}{4} + \frac{1}{\pi} \sum_{k=1}^{n-1} (n-k) \arcsin \lambda^k. \qquad (2.230)$$

Puis en vertu de (2.225) on a

$$\rho = \lim_{n\to\infty} \left[1 + \frac{4}{\pi} \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) \arcsin \lambda^k\right]^{-1}. \qquad (2.231)$$

En développant $\arcsin \lambda^k$ en série suivant les puissances de $\lambda h$ et en utilisant la formule de la progression arithmétique-géométrique, on obtient à partir de (2.231)

$$\rho = \left[1 + \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(2k)!\, \lambda^{2k+1}}{2^{2k} (k!)^2 (2k+1)(1-\lambda^{2k+1})}\right]^{-1}. \qquad (2.232)$$

Pour un bruit faiblement corrélé on a

$$\rho = 1 - 4\lambda/\pi + o(\lambda^2). \qquad (2.232')$$

---

*) Remarquons que pour la distribution normale du bruit (2.229) découle immédiatement de l'expression pour la fonction de corrélation à la sortie d'un limiteur parfait (voir tome I, formule (7.90)).

Pour un bruit laplacien de densité de probabilité $w_1(x) = \frac{1}{\sigma\sqrt{2}} e^{-\frac{\sqrt{2}|x|}{\sigma}}$ la fonction caractéristique est égale à

$$\theta_1(v) = \frac{1}{2}\int_{-\infty}^{\infty} e^{-|x|} e^{ivx}\, dx = \frac{1}{1+\sigma^2 v^2/2}$$

et en vertu de (2.216)

$$\Theta_{\eta_{m+k}}(v) = \frac{1+\lambda^{2k}\sigma^2 v^2/2}{1+\sigma^2 v^2/2} = \lambda^{2k} + (1-\lambda^{2k})/(1+\sigma^2 v^2/2),$$

d'où

$$w_{\eta_{m+k}}(z) = \lambda^{2k}\delta(z) + \frac{1-\lambda^{2k}}{\sigma\sqrt{2}} e^{-\frac{\sqrt{2}}{\sigma}|z|}. \tag{2.233}$$

La substitution de (2.233) dans (2.222) donne

$$B_u(k) + \frac{1}{4} = \frac{1}{\sigma\sqrt{2}}\int_0^{\infty} e^{-\sqrt{2}|x|/\sigma}\left(\lambda^{2k} + \frac{1-\lambda^{2k}}{2}\int_{-\lambda^k x\sqrt{2}/\sigma}^{\infty} e^{-|z|}\, dz\right) dx =$$

$$= \frac{1}{2}\int_0^{\infty} e^{-y}\left[\lambda^{2k} + \frac{1-\lambda^{2k}}{2}(1+1-e^{-\lambda^k y})\right] dy =$$

$$= \frac{1}{2}\int_0^{\infty} e^{-y}\left(\lambda^{2k} + 1 - \lambda^{2k} - \frac{1-\lambda^{2k}}{2} e^{-\lambda^k y}\right) dy =$$

$$= \frac{1}{2} - \frac{1-\lambda^{2k}}{4}\,\frac{1}{1+\lambda^k} = \frac{1}{4} + \frac{1}{4}\lambda^{2k},$$

et dans le cas envisagé

$$B_u(k) = \frac{1}{4}\lambda^{2k}. \tag{2.234}$$

La variance de la statistique de signe se trouve à partir de (2.221) et (2.234) :

$$M_2\left\{\sum_{i=1}^{n} u(x_i) \mid H\right\} = \frac{n}{4} + \frac{1}{2}\sum_{k=1}^{n-1}(n-k)\lambda^{2k} =$$

$$= \frac{n}{4}\left[\frac{1+\lambda}{1-\lambda} - \frac{2\lambda}{n}\,\frac{1-\lambda^{2k}}{(1-\lambda)^2}\right]. \tag{2.235}$$

Avec (2.235) et (2.225) on trouve le coefficient d'efficacité asymptotique relative de l'algorithme envisagé pour la distribution laplacienne du bruit corrélé

$$\rho = (1-\lambda)/(1+\lambda). \tag{2.236}$$

Pour un bruit faiblement corrélé

$$\rho = 1 - 2\lambda + o(\lambda). \tag{2.236'}$$

**2.7.5. Comparaison des efficacités des intégrations avec limitation et sans limitation.** Dans ce paragraphe nous allons étudier les algorithmes analogiques de détection d'un signal constant (positif) noyé dans un bruit markovien stationnaire additif de moyenne nulle et de fonction de corrélation égale à $\sigma^2 e^{-\lambda|\tau|}$. On sait (voir tome II, 3.5.4) que dans le cas d'un bruit markovien normal l'algorithme optimal (d'après le critère de Neyman-Pearson) de détection d'un signal constant consiste à comparer avec un seuil la statistique $[x(0)+x(T)]/(\lambda T)+\frac{1}{T}\int_0^T x(t)\,dt$, où $x(t)$ est la réalisation du processus observé sur l'intervalle $(0, T)$. Cet algorithme optimal est réalisé par un intégrateur parfait. Pour l'hypothèse $H$ la valeur moyenne de cette statistique est nulle, et la variance pour $\lambda T \gg 1$ est égale à

$$M_2\left\{\frac{x(0)+x(T)}{\lambda T}+\frac{1}{T}\int_0^T x(t)\,dt\right\}=\frac{2\sigma^2}{\lambda T}+o[(\lambda T)^{-1}]. \qquad (2.237)$$

Supposons maintenant que le bruit markovien ne suit pas la loi normale et qu'avant l'intégration on procède à une limitation parfaite de la réalisation observée. L'algorithme de détection consiste alors à comparer avec un seuil la statistique $\frac{1}{T}\int_0^T u[x(t)]\,dt$ qui est une statistique de signe limite lorsque le nombre $n$ d'échantillons discrets sur l'intervalle d'observation $T$ augmente indéfiniment. Dans ce cas (voir (2.217))

$$B_x(1)=\sigma^2 e^{-\lambda T/n}, \qquad (2.237')$$

d'où l'on trouve pour l'hypothèse $H$ et pour une distribution symétrique du bruit

$$m_1\left\{\frac{1}{T}\int_0^T u[x(t)]\,dt\,|H\right\}=\frac{1}{T}\int_0^T m_1\{u[x(t)]\,|H\}\,dt=\frac{1}{2}, \qquad (2.238)$$

$$M_2\left\{\frac{1}{T}\int_0^T u[x(t)]\,dt\,|H\right\}=\lim_{n\to\infty} M_2\left\{\frac{1}{n}\sum_{i=1}^{n} u(x_i)\,|H\right\}, \qquad (2.239)$$

avec $x_i = x(iT/n)$.

En vertu de (2.237) et (2.239), des résultats de 2.3.3 et du fait que la statistique $\frac{1}{T}\int_0^T u[x(t)]\,dt$ est normale pour $\lambda T \to \infty$, l'efficacité asymptotique relative du détecteur à limitateur et intégrateur par rapport au détecteur (optimal pour un bruit normal) n'utilisant

qu'un intégrateur se mesure par

$$\rho = \frac{2\sigma^2}{\lambda T} w_1^2(0)/M_2\left\{\frac{1}{T}\int_0^T u[x(t)]\,dt\,|\,H\right\}, \qquad (2.240)$$

où $w_1(x)$ est la densité de probabilité unidimensionnelle du bruit.

Pour un bruit à distribution normale (voir (2.229)) on a

$$M_2\left\{\frac{1}{T}\int_0^T u[x(t)]\,dt\,|\,H\right\} = \frac{1}{\pi}\int_0^1 (1-z)\arcsin e^{-\lambda T z}\,dz. \qquad (2.241)$$

Pour $\lambda T \gg 1$

$$M_2\left\{\frac{1}{T}\int_0^T u[x(t)]\,dt\,|\,H\right\} \sim \frac{1}{\pi}\int_0^1 \arcsin e^{-\lambda T z}\,dz = \frac{1}{\pi\lambda T}\sum_{n=0}^{\infty}\frac{a_n}{2n+1},$$

où $a_k$ sont les coefficients du développement en série de puissances de arc sin $x$ (voir (2.232)). Mais puisque

$$\int_0^1 \frac{\arcsin x}{x}\,dx = \sum_{n=0}^{\infty}\frac{a_n}{2n+1} = \frac{\pi}{2}\ln 2,$$

on a

$$M_2\left\{\frac{1}{T}\int_0^T u[x(t)]\,dt\,|\,H\right\} \sim \frac{\ln 2}{2\lambda T}. \qquad (2.242)$$

En substituant (2.242) dans (2.240) et compte tenu que pour une distribution normale $w_1^2(0) = 1/(2\pi\sigma^2)$, on trouve

$$\rho = 2/(\pi \ln 2) \approx 0{,}9. \qquad (2.243)$$

Cette valeur est voisine du coefficient d'efficacité asymptotique relative du détecteur de phase optimal par rapport à l'algorithme optimal (voir tome II, page 302).

Pour un bruit laplacien on a (voir (2.235))

$$M_2\left\{\frac{1}{T}\int_0^T u[x(t)]\,dt\,|\,H\right\} = \lim_{n\to\infty}\frac{1+e^{-\lambda T/n}}{4n(1-e^{-\lambda T/n})} = 1/(2\lambda T). \qquad (2.244)$$

En substituant (2.244) dans (2.240) et vu que pour la distribution de Laplace $w_1^2(0) = 1/(2\sigma^2)$, on trouve

$$\rho = 2, \qquad (2.245)$$

ce qui coïncide avec le coefficient d'efficacité asymptotique relative de l'algorithme de signe par rapport à l'algorithme linéaire pour un bruit laplacien (voir (2.72)).

**2.7.6. Algorithme signe-rang de détection d'un signal constant noyé dans un bruit corrélé.** Nous avons remarqué dans 2.4.2 que l'algorithme signe-rang (2.92) de détection d'un signal constant (positif) noyé dans un bruit normal indépendant additif ne cède pas beaucoup en efficacité asymptotique à l'algorithme linéaire (2.62), optimal pour ce type de bruits. Il est intéressant de comparer les efficacités asymptotiques des algorithmes (2.92) et (2.62) pour un bruit corrélé.

Remarquons tout d'abord que pour un bruit normal corrélé, ce n'est plus l'algorithme (2.62) de détection d'un signal constant $a > 0$ qui soit optimal suivant le critère de Neyman-Pearson mais l'algorithme de sommation pondérée (voir tome II, 5.3.1). Il est facile de trouver le coefficient d'efficacité asymptotique relative de l'algorithme linéaire (2.62) utilisé pour un bruit corrélé par rapport à ce même algorithme qu'on emploie pour un bruit indépendant. Il est évident que ce coefficient sera égal au rapport des variances pour un bruit indépendant et pour un bruit corrélé, soit

$$\rho = [\sum_{k=-\infty}^{\infty} R(k\tau)]^{-1}, \tag{2.246}$$

où $R(t)$ est le coefficient de corrélation d'un bruit stationnaire au sens large et $\tau$ l'intervalle de quantification.

Si, par exemple, $R(t) = e^{-\alpha|t|}$, en vertu de (2.246) on a

$$\rho = \left(1 + 2\sum_{k=1}^{\infty} e^{-\alpha\tau k}\right)^{-1} = \frac{1-e^{-\alpha\tau}}{1+e^{-\alpha\tau}}. \tag{2.246'}$$

Pour $\tau \to 0$ on a $\rho \to 0$ et pour $\tau \to \infty$, $\rho \to 1$. Le calcul de la variance d'une statistique signe-rang étant très ardu pour une distribution quelconque du bruit corrélé, nous nous limiterons au cas d'un bruit normal. On peut montrer (voir [23]) que le coefficient d'efficacité asymptotique relative de l'algorithme signe-rang de détection d'un signal constant (positif) noyé dans un bruit normal additif corrélé par rapport à l'algorithme linéaire est égal à

$$\rho = \sum_{k=-\infty}^{\infty} R(k\tau) \Big/ \sum_{k=-\infty}^{\infty} 2 \arcsin \frac{R(k\tau)}{2}. \tag{2.247}$$

Pour $\tau \to \infty$, on a $\rho \to (2\arcsin 1/2)^{-1} = 3/\pi$, ce qui coïncide avec le résultat cité dans 2.4.2.

Dans 2.6.2, nous avons obtenu l'algorithme (2.196), optimal pour la détection d'un signal constant noyé dans un bruit normal additif de valeur moyenne inconnue, en présence d'un échantillon d'apprentissage du bruit, lorsque l'échantillon observé et celui d'apprentissage sont indépendants. Le coefficient d'efficacité asymp-

totique relative de l'algorithme (2.196), pour un bruit normal corrélé, par rapport à cet algorithme utilisé avec un bruit indépendant, est donné par la formule (2.246). Si dans le cas envisagé on fait usage de l'algorithme à deux échantillons de Wilcoxon (voir 2.6.3), le coefficient d'efficacité asymptotique relative de cet alorgithme par rapport à l'algorithme (2.196) est donné par la formule (2.247).

**2.7.7. Algorithme de rang de détection d'un signal stochastique noyé dans un bruit corrélé.** Dans 2.5.1 nous avons obtenu l'algorithme (2.158), optimal pour la détection d'un signal normal noyé dans des bruits normaux additifs, utilisé dans un système à deux canaux, lorsque les échantillons observés étaient indépendants tant pour l'hypothèse $H$ (il n'y a pas de signal) que pour l'alternative $K$ (le signal est présent). Pour un bruit normal corrélé cet algorithme ne sera plus optimal. Le coefficient d'efficacité asymptotique relative de l'algorithme (2.158) pour un bruit corrélé par rapport à cet algorithme pour un bruit indépendant est, comme il est facile de le voir, égal à (comparer avec (2.246))

$$\rho = [\sum_{k=-\infty}^{\infty} R^2(k\tau)]^{-1}. \tag{2.248}$$

Si par exemple, le coefficient de corrélation d'un bruit stationnaire au sens large est égal à $R(t) = e^{-\alpha|t|}$, on tire de (2.248)

$$\rho = \left[1 + 2\sum_{k=1}^{\infty} e^{-2\alpha\tau k}\right]^{-1} = \frac{1-e^{-2\alpha\tau}}{1+e^{-2\alpha\tau}}. \tag{2.248'}$$

Supposons que pour la détection d'un signal stochastique on utilise l'algorithme de rang non paramétrique de Kendall (voir (2.192)). Dans le cas d'un bruit normal corrélé le coefficient d'efficacité asymptotique relative de cet algorithme par rapport à l'algorithme optimal (2.158) utilisé pour un bruit indépendant est égal à

$$\rho = \sum_{k=-\infty}^{\infty} R^2(k\tau) \Big/ \sum_{k=-\infty}^{\infty} 2\left[2\arcsin\frac{R(k\tau)}{2}\right]^2. \tag{2.249}$$

Pour $\tau \to \infty$ on a $\rho \to 9/(2\pi^2)$, ce qui coïncide avec le résultat évoqué dans 2.5.5.

## 2.8. PROBLÈMES

2.1. a) Montrer que la densité de probabilité simultanée d'un élément appartenant à un échantillon homogène indépendant et de son rang est égale à

$$w_2(x, R) = \frac{(n-1)!}{(R-1)!\,(n-R)!} F_1^{R-1}(x)\,[1-F_1(x)]^{n-R}\, w_1(x), \tag{1}$$

où $w_1(x)$, $F_1(x)$ sont la densité de probabilité et la fonction de répartition du processus dont on prend l'échantillon, et $n$ la taille de l'échantillon.

b) En utilisant la formule (1), montrer que le coefficient de corrélation d'un élément de l'échantillon et de son rang est

$$r_{x,R} = \left(\frac{n-1}{n+1}\frac{12}{\sigma^2}\right)^{1/2} \int_{-\infty}^{\infty} x w_1(x) \left[F_1(x) - \frac{1}{2}\right] dx, \tag{2}$$

où $\sigma^2$ est le moment centré deux de la distribution $F_1(x)$.

c) Calculer $r_{x,R}$ pour la distribution normale et la distribution uniforme et vérifiér que dans le premier cas $r_{x,R} = \left(\frac{3}{\pi}\frac{n-1}{n+1}\right)^{1/2}$ et dans le second $r_{x,R} = \left(\frac{n-1}{n+1}\right)^{1/2}$.

2.2. Soient $x_1, \ldots, x_n$ un échantillon indépendant de la distribution $F_1(x)$, et $x^{(i)}$ la $i$-ième statistique d'ordre de l'échantillon. Montrer que la fonction de répartition $F_{u_s}(z)$ de la variable aléatoire

$$u_s = F_1[x^{(s+r)}] - F_1[x^{(r)}] \tag{3}$$

ne dépend pas de $F_1(x)$ et peut s'écrire comme suit:

$$F_{u_s}(z) = P\{u_s \leqslant z\} = \mathrm{I}_z(s, n-s+1) =$$
$$= B_z(s, n-s+1)/B(s, n-s+1), \quad 0 \leqslant z \leqslant 1, \tag{4}$$

où

$$B_z(a, b) = \int_0^z t^{a-1}(1-t)^{b-1}\,dt, \quad B(a, b) =$$
$$= \int_0^1 t^{a-1}(1-t)^{b-1}\,dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

2.3. Soient $x_{c1}, \ldots x_{cn}$ et $x_{s1}, \ldots, x_{sn}$ des échantillons indépendants des composantes en quadrature du processus aléatoire à bande étroite observé qui est soit un bruit, soit un mélange d'un bruit additif et d'un signal sinusoïdal quasi déterministe à phase aléatoire. Généraliser au cas envisagé l'algorithme signe-rang de détection de signaux (voir (2.91)) et montrer que cet algorithme généralisé se réduit à la comparaison à un seuil de la statistique

$$\mathrm{S}_n(\mathrm{x}) = [\mathrm{S}_{nc}^{*2}(\mathrm{x}) + \mathrm{S}_{ns}^{*2}(\mathrm{x})]^{1/2}, \tag{5}$$

où

$$S^*_{nc}(x) = \max\left\{S_{nc}(x),\ \frac{n(n+1)}{2} - S_{nc}(x)\right\}; \tag{6}$$

$$S^*_{ns}(x) = \max\left\{S_{ns}(x),\ \frac{n(n+1)}{2} - S_{ns}(x)\right\}; \tag{7}$$

$$S_{nc}(x) = \sum_{i=1}^{n} R^+_{ci} u(x_{ci}); \tag{8}$$

$$S_{ns}(x) = \sum_{i=1}^{n} R^+_{si} u(x_{si}), \tag{9}$$

et $R^+_{ci}$, $R^+_{si}$ sont les rangs des modules des valeurs échantillonnées des composantes en quadrature.

2.4. Montrer que le coefficient d'efficacité asymptotique relative de l'algorithme signe-rang de détection d'un signal constant noyé dans un bruit additif logistique (voir (3.163)), par rapport à l'algorithme linéaire est égal à $\pi^2/9$.

2.5 [10]. Soient $x_1, \ldots, x_n$ et $y_1, \ldots, y_n$ des observations indépendantes à la sortie d'un système à deux canaux de détection d'un signal stochastique noyé dans des bruits additifs (voir 2.5.1). Pour la détection du signal, on utilise un algorithme prescrivant la comparaison à un seuil de la statistique

$$S_n(x, y) = \sum_{i=1}^{n} \operatorname{sgn}(x_i - x^*) \operatorname{sgn}(y_i - y^*), \tag{10}$$

où $x^*$ et $y^*$ sont les médianes échantillonnées. Montrer que le coefficient d'efficacité asymptotique relative de cet algorithme par rapport à l'algorithme (2.166) utilisant le coefficient de corrélation échantillonné est égal à $4/\pi^2$.

2.6. Montrer que pour la statistique (2.197), pour une distribution quelconque du bruit, la covariance entre le numérateur et l'expression dans l'accolade dans le dénominateur, pour l'hypothèse $H$, est égale à

$$\operatorname{cov} = M_3\left(\frac{1}{n_1} - \frac{1}{n_2}\right)\Big/(n_1 + n_2 - 2), \tag{11}$$

où $M_3$ est le moment centré trois de la distribution du bruit.

2.7. Soient $y_1, \ldots, y_n$ un échantillon d'apprentissage indépendant du bruit et $x_1, \ldots, x_n$ un échantillon indépendant servant à tester l'hypothèse selon laquelle le processus observé est un mélange d'un signal stochastique et d'un bruit additif (voir 2.6.1). Montrer que le coefficient d'efficacité asymptotique relative de l'algorithme mixte de détection d'un signal, prescrivant la comparaison à un

seuil de la statistique

$$S_n(x, y) = \sum_{i=0}^{r-1} S_m^{(i)}(x, y) = \sum_{i=0}^{r-1} \sum_{j=1}^{m} \sum_{k=1}^{m} \operatorname{sgn}(x_{im+j} - y_{im+k}), \quad (12)$$

par rapport à l'algorithme (2.196) est égal à

$$\rho = (3/\pi)\,[2m/(2m+1)], \quad (13)$$

où $m$ est le nombre de valeurs échantillonnées dans chacun des $r$ groupes (voir 2.6.4 et aussi [24]). Montrer que la formule (13) est également vraie pour le coefficient d'efficacité asymptotique relative de l'algorithme (12) par rapport à l'algorithme de Student (2.197).

CHAPITRE 3

# ALGORITHMES ASYMPTOTIQUEMENT OPTIMAUX DE DÉTECTION DES SIGNAUX

## 3.1. STRUCTURE DES ALGORITHMES ASYMPTOTIQUEMENT OPTIMAUX DE DÉTECTION DES SIGNAUX D'APRÈS DES OBSERVATIONS INDÉPENDANTES

**3.1.1. Développement asymptotique du logarithme du rapport de vraisemblance.** Soit $\mathbf{x} = (x_1, \ldots, x_n)$ un échantillon indépendant issu de la réalisation $x(t)$ du processus observé, on a $x_i = x(t_i)$, $i = 1, \ldots, n$. On vérifie l'hypothèse $H$ selon laquelle l'échantillon est homogène et appartient à la distribution du bruit $w_1(x; 0)$, contre l'alternative que l'élément $x_i$ de l'échantillon appartient au mélange d'un signal déterministe $\lambda_n s(t)$ et d'un bruit de distribution $w_1(x_i; \lambda_n s_i)$, par ailleurs $s_i = s(t_i)$, $i = 1, \ldots, n$. Comme nous l'avons noté dans 1.1.8, l'algorithme optimal de détection d'un signal déterministe noyé dans un bruit stationnaire indépendant consiste à réaliser la comparaison avec un seuil du rapport de vraisemblance

$$l(x_1, \ldots, x_n) = \prod_{i=1}^{n} w_1(x_i; \lambda_n s_i)/w_1(x_i; 0) \tag{3.1}$$

ou (ce qui est équivalent car la fonction logarithmique est monotone du logarithme du rapport de vraisemblance

$$\ln l(x_1, \ldots, x_n) = \sum_{i=1}^{n} \ln [w_1(x_i; \lambda_n s_i)/w_1(x_i; 0)], \tag{3.2}$$

qui pour $n$ quelconque est une statistique suffisante dans le problème de la vérification des hypothèses statistiques (1.1.8).

La synthèse des algorithmes asympotiquement optimaux de détection des signaux noyés dans des bruits (voir 1.2.8) est basée sur l'étude du développement asymptotique du logarithme du rapport de vraisemblance pour $n \to \infty$ et $\lambda_n \sqrt{n} = \gamma$, où $\gamma$ est une constante positive (voir (1.50)).

Avant de passer à cette étude nous allons formuler les conditions auxquelles doit satisfaire la densité de probabilité du mélange du signal et du bruit et que l'on rencontre habituellement dans les problèmes pratiques. Désignons par $\vartheta$ le paramètre de la fonction

$w_1(x;\ \lambda_n s)$

$$\vartheta = \lambda_n s = \gamma s/\sqrt{n},$$

où $s = s(t)$ est la valeur (normée) du signal à un instant quelconque. Exigeons que la densité de probabilité $w_1(x;\ \vartheta)$ 1) soit continue par rapport à $\vartheta$ au point $\vartheta = 0$ uniformément pour toutes les valeurs de $x$, 2) ne s'annule pas dans le domaine des valeurs échantillonnées et 3) admette le développement

$$w_1(x;\ \vartheta) = w_1(x;\ 0)\,[1 + \vartheta f(x) + \vartheta^2 \delta(x,\ \vartheta)], \tag{3.3}$$

où

$$f(x) = \frac{1}{w_1(x;\ 0)} \frac{\partial}{\partial \vartheta} w_1(x;\ \vartheta)|_{\vartheta=0}, \tag{3.4}$$

$$f(x) \not\equiv 0,$$

et

$$\mathrm{I}_f = m_1\{f^2(x)|H\} = \int_{-\infty}^{\infty} f^2(x)\, w_1(x;\ 0)\, dx < \infty, \tag{3.5}$$

$$m_1\{f(x)|H\} = \int_{-\infty}^{\infty} f(x)\, w_1(x;\ 0)\, dx = 0, \tag{3.6}$$

en outre, pour $\varepsilon > 0$ quelconque il existe toujours un $\vartheta_0$ tel que

$$\int_{-\infty}^{\infty} \delta^2(x;\ \vartheta)\, w_1(x;\ 0)\, dx < \varepsilon, \quad |\vartheta| < \vartheta_0. \tag{3.7}$$

Pour les conditions mentionnées le logarithme du rapport de vraisemblance admet le développement asymptotique suivant:

$$\ln l(x_1,\ \ldots,\ x_n) = \frac{\gamma}{\sqrt{n}} \sum_{i=1}^{n} s_i f(x_i) - \\ - \frac{\gamma^2}{2} \mathrm{I}_f \mathrm{W}_s + \alpha_n\left(x,\ \frac{\gamma}{\sqrt{n}} s\right), \tag{3.8}$$

où $\mathrm{W}_s$ est la puissance du signal *)

$$\mathrm{W}_s = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} s_i^2 \sim \lim_{T\to\infty} \frac{1}{T} \int_0^T s^2(t)\, dt, \quad \mathrm{W}_s < \infty. \tag{3.8'}$$

*) En introduisant les désignations $x_i = t_i/T$, $\Delta x_i = (t_{i+1} - t_i)/T = \frac{1}{n}$, on obtient pour $n \gg 1$

$$\frac{1}{n} \sum_{i=1}^{n} s^2(t_i) = \sum_{i=1}^{n} s^2(Tx_i)\, \Delta x_i \approx \int_0^1 s^2(Tx)\, dx = \frac{1}{T} \int_0^T s^2(t)\, dt.$$

Le terme résiduel $\alpha_n$ du développement (3.8) tend en probabilité vers zéro pour $n \to \infty$ tant pour l'hypothèse $H$, que pour l'alternative $K$.

Dans 3.1.2 nous avons donné une démonstration, bien que schématique, du développement de la forme (3.8) du logarithme du rapport de vraisemblance.

En examinant (3.8) on voit que les propriétés asymptotiques du logarithme du rapport de vraisemblance sont déterminées en premier lieu par le terme linéaire du développement, c'est-à-dire par la statistique

$$y_n(x_1, \ldots, x_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i f(x_i). \tag{3.9}$$

En vertu du théorème de la limite centrale, la statistique (3.9) est *asymptotiquement normale*, car on suppose que les termes de la somme sont indépendants et que pour $n$ suffisamment grand et pour des signaux réels le rapport de la variance d'un terme quelconque $s_i^2 \mathrm{I}_f/n$ à la variance de la somme $\mathrm{I}_f \mathrm{W}_s$ est inférieur à une grandeur quelconque aussi petite que l'on veut, c'est-à-dire que l'on a (cf. (2.119))

$$\lim_{n\to\infty} \max_{1\leq i\leq n} s_i^2 / \sum_{i=1}^{n} s_i^2 = 0. \tag{3.9'}$$

Trouvons les paramètres de la distribution asymptotiquement normale de la statistique $y_n$ (x) définie par (3.9). Pour l'hypothèse $H$ la moyenne de $y_n$ est nulle en vertu de (3.6), et la variance est

$$M_2\{y_n|H\} = M_2\{f|H\} \frac{1}{n} \sum_{i=1}^{n} s_i^2 \sim \mathrm{I}_f \mathrm{W}_s. \tag{3.10}$$

Pour l'alternative $K$ la moyenne de $y_n$ peut être calculée à partir du développement (3.3):

$$m_1\{y_n|K\} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i m_1\{f(x_i)|K\} =$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i \int_{-\infty}^{\infty} w_1(x_i;\ \gamma s_i/\sqrt{n})\, f(x_i)\, dx_i \sim$$

$$\sim \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i \int_{-\infty}^{\infty} \left[1 + \frac{\gamma s_i}{\sqrt{n}} f(x_i)\right] f(x_i)\, w_1(x_i;\ 0)\, dx_i =$$

$$= \gamma \mathrm{I}_f \frac{1}{n} \sum_{i=1}^{n} s_i^2;\quad m_1\{y_n|K\} \sim \gamma \mathrm{I}_f \mathrm{W}_s. \tag{3.11}$$

Des calculs plus longs permettent de voir que les variances de la statistique (3.9) pour l'alternative $K$ et pour l'hypothèse $H$ sont asymptotiquement égales (voir (3.10)). Par ailleurs, ce résultat découle immédiatement de certaines considérations générales, comme nous l'avons remarqué dans 3.1.2.

En comparant (3.8) et (3.9) on voit que la distribution du logarithme du rapport de vraisemblance est également asymptotiquement normale, et les paramètres de cette distribution sont $(-\gamma^2 I_f W_s/2, \gamma^2 I_f W_s)$ pour l'hypothèse $H$ et $(\gamma^2 I_f W_s/2, \gamma^2 I_f W_s)$ pour l'alternative $K$.

**3.1.2. Schéma de la démonstration du développement asymptotique.** Pour montrer que le développement (3.8) est vrai pour l'alternative on utilise la notion de suites asymptotiquement équivalentes de distributions (*contiguïté*) introduite par Le Cam [27].

Deux suites de distributions $w_n(\mathbf{x}; 0)$ et $w_n(\mathbf{x}; \boldsymbol{\vartheta})$ sont dites asymptotiquement équivalentes si la convergence en probabilité d'une statistique quelconque $t_n(\mathbf{x}) \to 0$, pour une suite de distributions, entraîne pour l'autre suite la convergence en probabilité de cette statistique vers zéro. Comme l'a montré Le Cam, la condition nécessaire et suffisante d'équivalence asymptotique de deux suites de distributions est

$$\int_{-\infty}^{\infty} e^z w^0(z)\, dz = 1, \tag{3.12}$$

où $w^0(z)$ est la fonction de répartition limite du logarithme du rapport de vraisemblance

$$\ln l(\mathbf{x}) = \ln w_n(\mathbf{x}; \boldsymbol{\vartheta})/w_n(\mathbf{x}; 0) \tag{3.12'}$$

dans l'hypothèse que $\boldsymbol{\vartheta} = 0$. Pour que les deux distributions $w_n(\mathbf{x}; \boldsymbol{\vartheta})$ et $w_n(\mathbf{x}; 0)$ soient asymptotiquement équivalentes il suffit que le logarithme du rapport de vraisemblance pour l'hypothèse nulle soit asymptotiquement normal de paramètres $(-M_2/2, M_2)$, où $M_2$ est la variance de la distribution asymptotique. Il est facile de voir qu'alors la condition (3.12) est vérifiée:

$$\int_{-\infty}^{\infty} e^z (2\pi M_2)^{-1/2} \exp\{-(z+M_2/2)^2/2M_2\}\, dz =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\{y\sqrt{M_2} - (y+\sqrt{M_2}/2)^2/2\}\, dy =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\{-(y-\sqrt{M_2}/2)^2/2\}\, dy = 1.$$

En vertu du théorème de Le Cam (voir [3], chap. VI, p. 1.4), si la distribution du logarithme du rapport de vraisemblance pour l'hypothèse nulle est asymptotiquement normale de paramètres $(-M_2/2, M_2)$, elle est également asymptotiquement normale pour une alternative « voisine » de paramètres $(M_2/2, M_2)$. Par exemple (voir 1.1.9), deux suites de distributions normales: $w_n$ (x; 0) d'un bruit corrélé et $w_n$ (x, λs) d'un mélange additif d'un signal déterministe et d'un bruit sont asymptotiquement équivalentes (pour $\lambda^2 T = \gamma^2 < \infty$, $T \to \infty$, où $T$ est la durée de l'observation).

Indiquons maintenant en traits généraux les raisonnements conduisant à la démonstration du développement (3.8). En utilisant (3.3) on peut présenter le logarithme du rapport de vraisemblance comme suit:

$$\ln l, (x_1 \ldots, x_n) = \sum_{i=1}^{n} \ln \left[1 + \frac{\gamma s_i}{\sqrt{n}} f(x_i)\right] + \alpha_n^{(1)} \left(\mathrm{x}, \frac{\gamma}{\sqrt{n}} \mathrm{s}\right);$$

en développant le logarithme en série de Taylor, on trouve

$$\ln l(x_1, \ldots, x_n) = \frac{\gamma}{\sqrt{n}} \sum_{i=1}^{n} s_i f(x_i) - \frac{\gamma^2}{2n} \sum_{i=1}^{n} s_i^2 f^2(x_i) + \\ + \alpha_n^{(1)} \left(\mathrm{x}, \frac{\gamma}{\sqrt{n}} \mathrm{s}\right) + \alpha_n^{(2)} \left(\mathrm{x}, \frac{\gamma}{\sqrt{n}} \mathrm{s}\right).$$

Tout d'abord on montre pour l'hypothèse $H$ la convergence en probabilité vers zéro des termes résiduels $\alpha_n^{(1)}$ et $\alpha_n^{(2)}$ pour $n \to \infty$. Puis pour l'hypothèse $H$ on montre que, d'après la loi des grands nombres, la densité de probabilité de la distribution du terme quadratique $\frac{\gamma^2}{2n} \sum_{i=1}^{n} s_i^2 f^2(x_i)$ devient une fonction delta, c'est-à-dire que ce terme converge en probabilité vers une constante égale à

$$m_1 \left\{\frac{\gamma^2}{2n} \sum_{i=1}^{n} s_i^2 f^2(x_i) | H\right\} \sim \gamma^2 \mathrm{I}_f \mathrm{W}_s/2.$$

Pour l'hypothèse $H$ la distribution de $\ln l(x_1, \ldots, x_n)$ est asymptotiquement normale de paramètres $(-\gamma^2 \mathrm{I}_f \mathrm{W}_s/2, \gamma^2 \mathrm{I}_f \mathrm{W}_s)$. Ainsi se vérifie la condition d'équivalence asymptotique des suites des distributions — pour l'hypothèse $H$: $w_n$ (x; 0) et pour l'alternative « voisine » $K$: $w_n$ (x; $\lambda_n$s). Par conséquent, en vertu de la définition de l'équivalence asymptotique de mesures probabilistes, pour l'alternative $K$ également, les termes résiduels $\alpha_n^{(1)}$ et $\alpha_n^{(2)}$ convergent en probabilité vers zéro pour $n \to \infty$ et $\lambda_n \sqrt{n} = \gamma$. et le terme quadratique converge alors vers la constante $\gamma^2 \mathrm{I}_f \mathrm{W}_s/2$,

Notons que dans le développement (3.8) on a

$$\alpha_n\left(\mathrm{x}, \frac{\gamma}{\sqrt{n}}\mathrm{s}\right) = \alpha_n^{(1)}\left(\mathrm{x}, \frac{\gamma}{\sqrt{n}}\mathrm{s}\right) + \alpha_n^{(2)}\left(\mathrm{x}, \frac{\gamma}{\sqrt{n}}\mathrm{s}\right) + \alpha_n^{(3)}\left(\mathrm{x}, \frac{\gamma}{\sqrt{n}}\mathrm{s}\right),$$

où

$$\alpha_n^{(3)}\left(\mathrm{x}, \frac{\gamma}{\sqrt{n}}\mathrm{s}\right) = \frac{\gamma^2}{2}\left[\mathrm{I}_f\mathrm{W}_s - \frac{1}{n}\sum_{i=1}^{n} s_i^2 f^2(x_i)\right]$$

pour $n \to \infty$ converge en probabilité vers zéro.

Souvent pour $n \gg 1$ la quantité $\frac{1}{n}\sum_{i=1}^{n} s_i^2 f^2(x_i)$ est remplacée par $\mathrm{I}_f\mathrm{W}_s$ (par exemple, lors du calcul de la grandeur du seuil dans l'algorithme de détection).

En vertu de l'équivalence asymptotique des mesures, la distribution du logarithme du rapport de vraisemblance pour l'alternative $K$ est asymptotiquement normale de paramètres $(\gamma^2\mathrm{I}_f\mathrm{W}_s/2, \gamma^2\mathrm{I}_f\mathrm{W}_s)$, et la distribution de la statistique $y_n$ dans (3.9) pour l'alternative $K$ est asymptotiquement normale de paramètres $(\gamma\mathrm{I}_f\mathrm{W}_s, \mathrm{I}_f\mathrm{W}_s)$ (voir (3.11)). Notons également que la « distance » entre les statistiques $y_n$ correspondant à des hypothèses « voisines » $H$ et $K$ tend vers

$$m_1\{y_n \mid K\} - m_1\{y_n \mid H\} \sim \gamma\mathrm{I}_f\mathrm{W}_s.$$

**3.1.3. Algorithme asymptotiquement optimal de détection d'un signal déterministe.** Il découle de (3.8) et (3.9) que pour $n \to \infty$ la statistique $y_n(\mathrm{x})$ ne diffère du rapport de vraisemblance, qui est une statistique suffisante dans le problème envisagé de la détection du signal, que par un facteur $\gamma$ et un terme additif $\gamma^2\mathrm{I}_f\mathrm{W}_s/2$. Donc on peut affirmer que la statistique $y_n(\mathrm{x})$ est *asymptotiquement suffisante*. La règle asymptotiquement normale de détection d'un signal déterministe $\lambda s(t)$ noyé dans des bruits indépendants de distribution $w_1(x; 0)$ peut se formuler comme suit *): pour $n \gg 1$ on prend la décision que le signal est présent, si

$$y_n(\mathrm{x}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i f(x_i) \geqslant C, \tag{3.13}$$

et la décision qu'il n'y a pas de signal si c'est l'inégalité inverse de (3.13) qui est vérifiée.

---

*) Les conditions de « proximité » de l'hypothèse $H$ et de l'alternative $K$ restreignent quelque peu la classe des interactions admissibles du signal et du bruit. Notamment, il faut que pour $\lambda \to 0$ le signal disparaisse et le bruit reste (par exemple, $x(t) = \lambda s(t)\,\xi(t)$ est inadmissible, et $x(t) = [1 + \lambda s(t)]\,\xi(t)$ est tout à fait acceptable).

En utilisant la notion de fonction de décision (voir 1.1.6) on peut écrire comme suit l'algorithme de détection (3.13):

$$\Phi(\mathbf{x}) = 1 \text{ (le signal est présent) lorsque } y_n(\mathbf{x}) \geqslant C,$$
$$\Phi(\mathbf{x}) = 0 \text{ (il n'y a pas de signal) lorsque } y_n(\mathbf{x}) < C. \quad (3.13')$$

Par la suite nous utiliserons parfois cette représentation de l'algorithme de détection par les fonctions de décision.

La statistique $y_n$ étant asymptotiquement normale et compte tenu des valeurs des paramètres de la distribution limite pour l'hypothèse $H$, on trouve la grandeur du seuil $C$ à partir de la relation suivante pour une probabilité donnée $\alpha$ de fausse alarme:

$$P\{y_n \geqslant C \mid H\} = \frac{1}{\sqrt{2\pi}} \int\limits_{C/\sqrt{\mathrm{I}_f \mathrm{W}_s}}^{\infty} e^{-z^2/2}\, dz = \alpha, \quad (3.14)$$

d'où

$$C = x_\alpha \sqrt{\mathrm{I}_f \mathrm{W}_s}, \quad (3.15)$$

où $x_\alpha$ est l'$\alpha$-quantile de la loi normale.

Connaissant $C$, on trouve la probabilité de détection correcte en se servant des valeurs des paramètres de la distribution normale limite de la statistique $y_n$ pour l'alternative, pour une probabilité donnée $\beta$ de perte de signal:

$$P\{y_n \geqslant C \mid K\} = \frac{1}{\sqrt{2\pi}} \int\limits_{\frac{C-\gamma \mathrm{I}_f \mathrm{W}_s}{\sqrt{\mathrm{I}_f \mathrm{W}_s}}}^{\infty} e^{-z^2/2}\, dz = 1 - \beta \quad (3.16)$$

ou

$$x_{1-\beta} = \frac{C - \gamma \mathrm{I}_f \mathrm{W}_s}{\sqrt{\mathrm{I}_f \mathrm{W}_s}}, \quad (3.17)$$

où $x_{1-\beta}$ est l'$(1-\beta)$-quantile de la loi normale.

En substituant dans (3.17) la grandeur $C$ de (3.15) on trouve

$$x_{1-\beta} = x_\alpha - \gamma \sqrt{\mathrm{I}_f \mathrm{W}_s}, \quad (3.18)$$

d'où l'on trouve la grandeur $\gamma$ correspondant à des probabilités données de fausse alarme $\alpha$ et de perte de signal $\beta$

$$\gamma = (x_\alpha - x_{1-\beta}) / \sqrt{\mathrm{I}_f \mathrm{W}_s}, \quad \alpha < 1 - \beta. \quad (3.18')$$

Remarquons que pour des $\alpha$, $\beta$, $\mathrm{I}_f$, $\mathrm{W}_s$ donnés l'expression (3.18') donne également la valeur du seuil de l'amplitude du signal, soit

$$\lambda_n = \frac{1}{\sqrt{n}} \gamma(\alpha, \beta, \mathrm{I}_f, \mathrm{W}_s).$$

La caractéristique de fonctionnement (3.18) de l'algorithme asymptotiquement optimal (3.13) de détection d'un signal déterministe

pour $\gamma$ donné et $n \to \infty$ tend vers la caractéristique de fonctionnement de l'algorithme de détection optimal (d'après le critère de Neyman-Pearson) pour la même constante $\gamma$.

L'algorithme asymptotiquement optimal (3.13) de détection d'un signal déterministe admet une interprétation simple (fig. 3.1). Le dispositif de détection se compose dans ce cas de deux blocs: le premier est un convertisseur non linéaire non inertiel qui assure la transformation de l'échantillon observé $x$ en la grandeur $f(x)$,

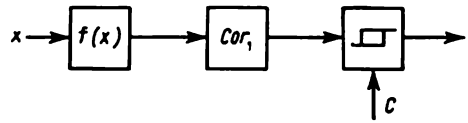



Fig. 3.1. Schéma fonctionnel du détecteur asymptotiquement optimal d'un signal déterministe (observations indépendantes).

et le second — un corrélateur effectuant la multiplication de la sortie du premier bloc par les valeurs du signal (en provenance d'un oscillateur local) et la sommation. Le corrélateur est suivi d'un dispositif à seuil. La décision que le signal est présent est prise lorsque ce seuil est dépassé.

**3.1.4. Caractéristique du convertisseur linéaire pour un bruit additif.** Pour un bruit additif on a $w_1(x; \vartheta) = w_1(x - \vartheta)$ et en vertu de (3.4)

$$f(x) = -w_1'(x;\ 0)/w_1(x;\ 0) = -[\ln w_1(x;\ 0)]'. \tag{3.19}$$

Dans ce cas la quantité d'information d'après Fisher quant à la distribution du bruit est (voir (3.5))

$$I_f = \int_{-\infty}^{\infty} [w_1'(x;\ 0)/w_1(x;\ 0)]^2 w_1(x;\ 0)\, dx.$$

Le produit $I_f W_s$ peut être interprété comme le rapport de la puissance du signal à la puissance du bruit.

Pour un bruit normal additif de moyenne nulle et de variance $\sigma^2$ la densité de probabilité est

$$w_1(x;\ 0) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}, \tag{3.20}$$

et par conséquent

$$f(x) = -\left(-\frac{x^2}{2\sigma^2}\right)' = \frac{x}{\sigma^2}. \tag{3.21}$$

Dans ce cas l'algorithme (3.13) coïncide avec l'algorithme optimal linéaire bien connu (algorithme corrélationnel). On prend la

décision que le signal est présent si (voir (3.13) et (3.21))

$$\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i s_i}{\sigma^2} \geqslant C. \tag{3.21'}$$

La quantité d'information d'après Fisher pour un bruit normal est égale à

$$I_f = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \frac{x^2}{\sigma^4} e^{-x^2/2}\, dx = \frac{1}{\sigma^2}, \tag{3.22}$$

et la grandeur $I_f W_s = W_s/\sigma^2$ n'est autre que le rapport habituel signal/bruit.

Dans le cas d'un bruit laplacien additif de puissance moyenne $\sigma^2$ la densité de probabilité est

$$w_1(x;\, 0) = \frac{1}{\sqrt{2\sigma^2}} e^{-\sqrt{\frac{2}{\sigma^2}}|x|}, \tag{3.23}$$

et la caractéristique du convertisseur non linéaire est de la forme

$$f(x) = \sqrt{2/\sigma^2}\ \mathrm{sgn}\, x, \tag{3.24}$$

où l'on reconnaît la caractéristique du limiteur parfait. Ainsi, si le corrélateur est précédé d'un limiteur parfait, on obtient un détecteur de signaux déterministe asymptotiquement optimal pour un bruit laplacien.

En portant (3.24) dans (3.13), on trouve

$$\sqrt{\frac{2}{n}} \sum_{i=1}^{n} \frac{s_i\, \mathrm{sgn}\, x_i}{\sigma} \geqslant C, \tag{3.25}$$

cette expression ne différant de (2.80″) que par un facteur constant. Par conséquent, le détecteur de signe est asymptotiquement optimal pour la détection d'un signal déterministe noyé dans un bruit laplacien additif. Autrement dit, le système limiteur parfait — corrélateur, utilisé parfois comme un dispositif non paramétrique de détection de signaux en présence d'un bruit de puissance inconnue pour la stabilisation du niveau de fausses alarmes, n'est asymptotiquement optimal que pour un bruit laplacien.

La quantité d'information d'après Fisher pour un bruit laplacien est égale à

$$I_f = \frac{1}{\sigma^2}\sqrt{\frac{2}{\sigma^2}} \int_{-\infty}^{\infty} e^{-\sqrt{\frac{2}{\sigma^2}}|x|}\, dx = \frac{2}{\sigma^2}, \tag{3.26}$$

d'où l'on trouve le produit

$$I_f W_s = 2W_s/\sigma^2 \tag{3.26'}$$

qui pour un bruit laplacien est le double du rapport signal/bruit. C'est ce qui explique le fait que l'algorithme de signe utilisé pour la détection d'un signal constant noyé dans un bruit laplacien additif est deux fois (asymptotiquement) plus efficace que l'algorithme linéaire (voir (2.72')).

**3.1.5. Stabilité de l'algorithme asymptotiquement optimal de détection d'un signal déterministe.** Revenons à l'algorithme asymptotiquement optimal (3.13) de détection d'un signal déterministe $\lambda s(t)$ noyé dans un bruit à valeurs indépendantes. Supposons que la distribution du bruit a changé et se caractérise par la densité $u_1(x;\ 0)$ au lieu de la densité de probabilité $w_1(x;\ 0)$ pour laquelle l'algorithme (3.13) était asymptotiquement optimal. Il est évident cependant que la statistique (3.9), mise à la base de l'algorithme (3.13), est pour un échantillon indépendant $x_1, \ldots, x_n$ asymptotiquement normale également dans le cas où elle appartient à la distribution $u_1(x;\ 0)$. Seuls les paramètres de la distribution asymptotiquement normale changent.

Calculons la valeur moyenne et la variance de la statistique (3.9) pour l'hypothèse $H$, lorsque l'échantillon du bruit appartient à la distribution $u_1(x;\ 0)$:

$$m_1\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i f(x_i)\,|\,H,\ u\right\} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i \int_{-\infty}^{\infty} f(x)\,u_1(x;\ 0)\,dx, \qquad (3.27)$$

$$M_2\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i f(x_i)\,|\,H,\ u\right\} =$$

$$= \left[\int_{-\infty}^{\infty} f^2(x)\,u_1(x;\ 0)\,dx - \left(\int_{-\infty}^{\infty} f(x)\,u_1(x;\ 0)\,dx\right)^2\right] \mathrm{W}_{sn}, \qquad (3.28)$$

où $\mathrm{W}_{sn} = \frac{1}{n}\sum_{i=1}^{n} s_i^2$.

Pour l'alternative $K$, lorsque l'échantillon $x_i$ appartient à la distribution $u_1(x_i;\ \lambda s_i)$, la valeur moyenne de la statistique (3.9) est égale à

$$m_1\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i f(x_i)\,|\,K,\ u\right\} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i \int_{-\infty}^{\infty} f(x)\,u_1(x;\ \lambda s_i)\,dx$$

et pour $\lambda\sqrt{n} = \gamma$, $n \to \infty$

$$u_1(x;\ \lambda s_i) = u_1(x;\ 0) + \lambda s_i g(x)\,u_1(x;\ 0) + o(1/\sqrt{n}),$$

où

$$g(x)=\frac{1}{u_1(x;\,0)}\frac{\partial}{\partial\vartheta}u_1(x;\,\vartheta)|_{\vartheta=0}, \tag{3.29}$$

on obtient alors

$$m_1\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}s_i f(x_i)\,|\,K,\,u\right\}\sim\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}s_i\right)\int_{-\infty}^{\infty}f(x)\,u_1(x;\,0)\,dx+$$

$$+\gamma W_{sn}\int_{-\infty}^{\infty}f(x)\,g(x)\,u_1(x;\,0)\,dx. \tag{3.30}$$

La variance de la statistique (3.9) pour l'alternative $K$, avec les mêmes conditions, est

$$M_2\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}s_i f(x_i)\,|\,K,\,u\right\}\sim M_2\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}s_i f(x_i)\,|\,H,\,u\right\},$$

c'est-à-dire qu'elle est donnée par (3.28).

Si au lieu de la statistique (3.13) on utilise la statistique centrée

$$y_n(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(s_i-\bar{s})\,f(x_i), \tag{3.31}$$

où

$$\bar{s}=\frac{1}{n}\sum_{i=1}^{n}s_i, \tag{3.32}$$

on a $\sum_{i=1}^{n}(s_i-\bar{s})=0$ et à partir de (3.27), (3.28), (3.30) on obtient

$$m_1\{y_n\,|\,H,\,u\}=0,$$

$$m_1\{y_n\,|\,K,\,u\}=\gamma W_{sn}\int_{-\infty}^{\infty}f(x)\,g(x)\,u_1(x;\,0)\,dx, \tag{3.33}$$

$$M_2\{y_n\,|\,H,\,u\}\sim M_2\{y_n\,|\,K,\,u\}\sim$$

$$\sim W_s\left[\int_{-\infty}^{\infty}f^2(x)\,u_1(x;\,0)\,dx-\left(\int_{-\infty}^{\infty}f(x)\,u_1(x;\,0)\,dx\right)^2\right]. \tag{3.33'}$$

L'algorithme asymptotiquement optimal de détection d'un signal déterministe noyé dans un bruit de distribution $u_1(x;\,0)$ s'obtient à partir de (3.13) en remplaçant la caractéristique $f(x)$ du convertisseur non linéaire d'entrée par la caractéristique $g(x)$

définie par (3.29), c'est-à-dire

$$y_{n,u}(\mathrm{x}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i g(x_i). \tag{3.34}$$

Les paramètres de la statistique asymptotiquement normale (3.34) sont :

— pour l'hypothèse $H$ (voir (3.6) et (3.10))

$$m_1\{y_{n,u}(\mathrm{x}) \mid H\} = \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i\right) \int_{-\infty}^{\infty} g(x)\, u_1(x;0)\, dx = 0, \tag{3.35}$$

$$M_2\{y_{n,u}(\mathrm{x}) \mid H\} \sim \mathrm{W}_s \int_{-\infty}^{\infty} g^2(x)\, u_1(x;0)\, dx = \mathrm{I}_g \mathrm{W}_s, \tag{3.36}$$

— pour l'hypothèse $K$ (voir (3.11))

$$m_1\{y_{n,u}(\mathrm{x}) \mid K\} \sim \gamma \mathrm{I}_g \mathrm{W}_s, \tag{3.37}$$

$$M_2\{y_{n,u}(\mathrm{x}) \mid K\} \sim \mathrm{I}_g \mathrm{W}_s, \tag{3.38}$$

où $\mathrm{I}_g$ est la quantité d'information d'après Fisher sur la distribution du bruit $u_1(x;0)$.

Pour caractériser la stabilité de l'algorithme (3.13), servons-nous de la formule (1.77) et trouvons son coefficient d'efficacité asymptotique relative en présence d'un bruit de distribution $u_1(x;0)$ par rapport à l'algorithme (3.34) asymptotiquement optimal pour ce bruit « étranger ». Remarquons que la formule (1.77) est vraie pour les statistiques centrées pour l'hypothèse $H$. La statistique (3.34) pour un bruit de distribution $u_1(x;0)$ est centrée pour l'hypothèse $H$ (voir (3.35)) ; quant à la statistique (3.13) pour un bruit « étranger », il faut la centrer, c'est-à-dire passer à la statistique (3.31).

Avec les notations de la formule (1.77) on obtient dans le cas envisagé (voir (3.33), (3.33′), (3.37), (3.38)), ayant en vue que la statistique (3.31) est centrée :

$$a = \gamma \mathrm{I}_g \mathrm{W}_s, \quad \sigma^2 = \mathrm{I}_g \mathrm{W}_s, \tag{3.39}$$

$$a_1 = \gamma \mathrm{W}_s \int_{-\infty}^{\infty} f(x)\, g(x)\, u_1(x;0)\, dx, \tag{3.40}$$

$$\sigma_1^2 = \mathrm{W}_s \left[ \int_{-\infty}^{\infty} f^2(x)\, u_1(x;0)\, dx - \left( \int_{-\infty}^{\infty} f(x)\, u_1(x;0)\, dx \right)^2 \right]. \tag{3.41}$$

En substituant (3.39) à (3.41) dans (1.77), on trouve le coefficient cherché d'efficacité asymptotique relative caractérisant la

stabilité de l'algorithme (3.31):

$$\rho = \left( \int_{-\infty}^{\infty} f(x)\, g(x)\, u_1(x;0)\, dx \right)^2 \left\{ \int_{-\infty}^{\infty} g^2(x)\, u_1(x;0)\, dx \times \right.$$

$$\left. \times \left[ \int_{-\infty}^{\infty} f^2(x)\, u_1(x;0)\, dx - \left( \int_{-\infty}^{\infty} f(x)\, u_1(x;0)\, dx \right)^2 \right] \right\}^{-1}. \qquad (3.42)$$

Si les distributions $w_1(x;0)$ et $u_1(x;0)$ sont symétriques par rapport à l'origine, on a $\int_{-\infty}^{\infty} f(x)\, u_1(x;0)\, dx = 0$ et il n'est donc plus nécessaire de centrer la statistique pour un bruit « étranger ». La formule (3.42) prend alors un aspect plus simple:

$$\rho = \frac{\left( \int_{-\infty}^{\infty} f(x)\, g(x)\, u_1(x;0)\, dx \right)^2}{\left[ \int_{-\infty}^{\infty} g^2(x)\, u_1(x;0)\, dx \int_{-\infty}^{\infty} f^2(x)\, u_1(x;0)\, dx \right]}. \qquad (3.43)$$

L'expression (3.43) est le carré du coefficient de corrélation des statistiques $f(x)$ et $g(x)$ lorsque $x$ suit la loi $u_1(x;0)$. Par conséquent, $\rho \leqslant 1$, le cas $\rho = 1$ ayant lieu seulement si $f(x) \equiv g(x)$.

Voyons deux exemples illustrant la formule (3.43). Soit $f(x) = \frac{x}{\sigma^2}$. Comme nous l'avons dit plus haut (voir (3.21)), dans ce cas l'algorithme (3.13) coïncide avec l'algorithme linéaire, optimal pour un bruit additif, lorsque

$$w_1(x;0) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-x^2/2\sigma^2}.$$

Supposons que l'on utilise un algorithme linéaire pour la détection d'un signal déterministe noyé dans un bruit additif distribué suivant une loi de Laplace de même variance $\sigma^2$ (voir (3.23))

$$u_1(x;0) = \frac{1}{\sqrt{2\sigma^2}}\, e^{-\sqrt{\frac{2}{\sigma^2}}|x|}. \qquad (3.44)$$

En vertu de (3.34), on a dans ce cas (voir (3.24))

$$g(x) = \sqrt{2/\sigma^2}\, \operatorname{sgn} x. \qquad (3.44')$$

Pour un bruit laplacien, le coefficient d'efficacité asymptotique relative d'un algorithme linéaire par rapport à un algorithme asymp-

totiquement optimal pour ce bruit est en vertu de (3.43) égal à

$$\rho = \left(\frac{1}{\sigma^4}\int_{-\infty}^{\infty} xe^{-\sqrt{\frac{2}{\sigma^2}}|x|}\operatorname{sgn} x\,dx\right)^2 \times$$

$$\times\left[\frac{1}{\sigma^2}\int_{-\infty}^{\infty} e^{-\sqrt{\frac{2}{\sigma^2}}|x|}dx\int_{-\infty}^{\infty} x^2e^{-\sqrt{\frac{2}{\sigma^2}}|x|}dx\right]^{-1} = \frac{1}{2}, \qquad (3.45)$$

ce qui est en accord avec (2.72').

Ainsi, l'efficacité asymptotique d'un algorithme linéaire, optimal pour un bruit normal, diminue de moitié lorsqu'on l'utilise pour la détection d'un signal noyé dans un bruit laplacien (« étranger »).

Soit $f(x) = \sqrt{2/\sigma^2}\operatorname{sgn} x$. Cette fonction que réalise un convertisseur non linéaire d'entrée correspond à un algorithme optimal de détection d'un signal déterministe noyé dans un bruit additif laplacien (voir (3.24)). Supposons que cet algorithme soit utilisé pour un bruit additif normal de même variance. On a alors $g(x) = x/\sigma^2$, et à partir de (3.43) on trouve

$$\rho = \left(\frac{1}{\sigma^4\sqrt{\pi}}\int_{-\infty}^{\infty} x\operatorname{sgn} xe^{-x^2/2\sigma^2}dx\right)^2 \times$$

$$\times\left[\frac{1}{\sigma^6 2\pi}\int_{-\infty}^{\infty} x^2e^{x^2/2\sigma^2}dx\int_{-\infty}^{\infty} e^{-x^2/2\sigma^2}dx\right]^{-1} = \frac{2}{\pi}, \qquad (3.46)$$

ce qui, comme on pouvait s'y attendre, coïncide avec (2.70).

**3.1.6. Algorithme asymptotiquement optimal de détection d'un signal quasi déterministe.** Considérons maintenant un signal d'une forme plus générale (voir (1.9))

$$s(t) = \sum_{i=1}^{m}\vartheta_i\varphi_i(t) = \boldsymbol{\vartheta}'\boldsymbol{\varphi}(t), \qquad (3.47)$$

où $\boldsymbol{\vartheta}' = (\vartheta_1, \ldots, \vartheta_m)$ est un ensemble de paramètres aléatoires liés et $\boldsymbol{\varphi}(t) = (\varphi_1(t), \ldots, \varphi_m(t))$ sont des fonctions données. Désignons par $\boldsymbol{\varphi}_i$ le vecteur (à $m$ dimensions) obtenu à partir de la fonction vectorielle $\boldsymbol{\varphi}(t)$ pour $t = t_i$, i.e.

$$\boldsymbol{\varphi}_i = \boldsymbol{\varphi}(t_i) = (\varphi_1(t_i), \ldots, \varphi_m(t_i)), \quad i = 1, \ldots, n. \qquad (3.48)$$

En vertu de (3.47) on a alors

$$s_i = s(t_i) = \boldsymbol{\vartheta}'\boldsymbol{\varphi}(t_i) = \boldsymbol{\vartheta}'\boldsymbol{\varphi}_i. \qquad (3.48')$$

En substituant (3.48') dans (3.8), on trouve le développement asymptotique du logarithme du rapport de vraisemblance (condi-

tionnel, c'est-à-dire pour $\vartheta$ donné) pour le signal envisagé, soit $\ln l(x_1, \ldots, x_n \mid \vartheta) =$

$$= \frac{\gamma\vartheta'}{\sqrt{n}} \sum_{i=1}^{n} \varphi_i f(x_i) - \frac{\gamma^2}{2} \mathrm{I}_f \vartheta' \mathbf{A} \vartheta + \alpha_n (\mathbf{x}, \gamma \mathbf{s}/\sqrt{n}), \qquad (3.49)$$

où $\mathbf{A}$ est la matrice $m \times m$ dont les éléments $a_{jk}$ sont

$$a_{jk} = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \varphi_j(t_i) \varphi_k(t_i) \sim$$

$$\sim \lim_{T \to \infty} \frac{1}{T} \int_0^T \varphi_j(t) \varphi_k(t)\, dt, \quad j, k = 1, \ldots, m. \qquad (3.50)$$

Tout comme dans (3.8) le terme résiduel $\alpha_n$ dans (3.49) converge en probabilité pour $n \to \infty$ vers zéro, tant pour l'hypothèse que pour l'alternative.

Considérons la statistique vectorielle à $m$ dimensions

$$\mathbf{y}_n(\mathbf{x}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi_i f(x_i), \qquad (3.51)$$

qui correspond au terme linéaire dans le développement (3.49).

Si pour $k = 1, \ldots, m$ quelconque

$$\lim_{n \to \infty} \max_{1 \leqslant i \leqslant n} \varphi_k^2(t_i) \Big/ \sum_{i=1}^{n} \varphi_k^2(t_i) = 0,$$

chacune des composantes de la statistique vectorielle $\mathbf{y}_n(\mathbf{x})$

$$y_n^{(k)}(\mathbf{x}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi_k(t_i) f[x(t_i)]$$

est en vertu du théorème de la limite centrale *asymptotiquement normale*, de paramètres $(0, \mathrm{I}_f a_{kk})$ pour l'hypothèse $H$ et de paramètres $(\gamma \mathrm{I}_f \sum_{j=1}^{m} \vartheta_j a_{jk}, \mathrm{I}_f a_{kk})$ pour l'alternative $K$.

Comme pour $n \to \infty$

$$m_1 \{y_n^{(j)}(\mathbf{x})\, y_n^{(k)}(\mathbf{x})\} = \mathrm{I}_f a_{jk}$$

et la matrice $\mathrm{I}_f \mathbf{A}$ est définie positive, en vertu du théorème multidimensionnel de la limite centrale la statistique (3.51) est *asymptotiquement normale*, la matrice de corrélation de la distribution limite normale à $m$ dimensions étant égale à $\mathrm{I}_f \mathbf{A}$, et le vecteur des moyennes est égal à zéro si $\mathbf{x}$ est un échantillon du bruit, et à $\gamma \mathrm{I}_f \mathbf{A} \vartheta$

si $\mathbf{x}$ est un échantillon du mélange du signal et du bruit. Par conséquent, le logarithme du rapport de vraisemblance conditionnel (3.49) est également réparti suivant une loi asymptotiquement normale de paramètres $\left(-\frac{\gamma^2}{2} I_f \vartheta' A \vartheta,\ \gamma^2 I_f \vartheta' A \vartheta\right)$ pour l'hypothèse $H$ et de paramètres $\left(\frac{\gamma^2}{2} I_f \vartheta' A \vartheta,\ \gamma^2 I_f \vartheta' A \vartheta\right)$ pour l'alternative $K$.

Si les paramètres $\vartheta = (\vartheta_1, \ldots, \vartheta_m)$ du signal sont aléatoires et que l'on connaisse la distribution conjointe a priori $w_m(\vartheta)$, le développement asymptotique moyenné du rapport de vraisemblance (moyenne sur $\vartheta$ du rapport de vraisemblance conditionnel) se trouve à partir de (3.49) (rejetant le terme résiduel et utilisant la désignation de la statistique (3.51)) *)

$$\Lambda[y_n(\mathbf{x})] = \int_{\Theta} l(\mathbf{x}/\vartheta)\, w_m(\vartheta)\, d\vartheta =$$

$$= \int_{\Theta} w_m(\vartheta) \exp\left(\gamma \vartheta' y_n(\mathbf{x}) - \frac{\gamma^2}{2} I_f \vartheta' A \vartheta\right) d\vartheta. \qquad (3.52)$$

Il est intéressant de comparer la formule (3.52) avec (1.32). Il y a lieu de remarquer que la distribution de la statistique $\Lambda[y_n(\mathbf{x})]$ obtenue après la prise de moyenne sur $\vartheta$ n'est dans le cas général asymptotiquement normale.

La règle asymptotiquement optimale de détection du signal quasi déterministe (3.47) peut maintenant se formuler comme suit :

$$\Phi[y_n(\mathbf{x}) = 1 \text{ (le signal est présent) si } \Lambda[y_n(\mathbf{x})] \geqslant C, \qquad (3.53)$$
$$\Phi[y_n(\mathbf{x})] = 0 \text{ (le signal est absent) si } \Lambda[y_n(\mathbf{x})] < C,$$

où $\Lambda[y_n(\mathbf{x})]$ est donné par (3.52), le seuil $C$ et la constante $\gamma$ se trouvent pour des probabilités données de fausse alarme $\alpha$ et de détection correcte $1 - \beta$ à partir des relations

$$P\{\Lambda(y_n) \geqslant C \mid H\} = m_1\{\Phi[y_n(\mathbf{x})] \mid H\} = \alpha, \qquad (3.54)$$
$$P\{\Lambda(y_n) \geqslant C \mid K\} = m_1\{\Phi[y_n(\mathbf{x})] \mid K\} = 1 - \beta. \qquad (3.55)$$

Portons (3.53) dans (3.54), (3.55) ; compte tenu de ce que la distribution de la statistique $y_n(\mathbf{x})$ est asymptotiquement normale et des valeurs des paramètres de cette distribution pour l'hypothèse $H$ et l'alternative $K$, on peut écrire sous forme explicite les expres-

*) Notons que la moyenne est prise non pas sur le paramètre informationnel $\lambda$ du signal (pour les problèmes de la détection), mais sur le paramètre perturbateur (voir par exemple 3.1.10). On aurait pu introduire également les paramètres perturbateurs du bruit $\theta$, en se donnant la distribution simultanée du signal et du bruit sous la forme $w(x; \lambda's(t), \theta)$ et en prenant la moyenne sur tous les paramètres perturbateurs.

sions de $\alpha$ et $1-\beta$, après avoir pris la moyenne sur $\mathbf{y}_n$, soit :

$$(2\pi)^{-m/2}(\det \mathrm{I}_f\mathbf{A})^{-1/2}\int\limits_{\mathbf{Y}_m}\int\limits_{\Theta}\Phi(\mathbf{y})\exp\left\{-\frac{1}{2\mathrm{I}_f}(\mathbf{y}-\right.$$

$$\left.-\gamma\mathrm{I}_f\mathbf{A}\boldsymbol{\vartheta})'\mathbf{A}^{-1}(\mathbf{y}-\gamma\mathrm{I}_f\mathbf{A}\boldsymbol{\vartheta})\,w_m(\boldsymbol{\vartheta})\,d\boldsymbol{\vartheta}\,d\mathbf{y}=\begin{cases}\alpha, & \gamma=0,\\ 1-\beta, & \gamma>0.\end{cases} \tag{3.56}$$

La formule (3.56) est un système de deux équations par rapport à la constante $\gamma$ et le seuil $C$, celui-ci apparaît dans (3.56) d'une manière implicite, par la fonction de décision $\Phi(y)$, qui dans l'espace euclidien des statistiques $\mathbf{y}_n$ délimite lors de l'intégration le sous-espace vérifiant l'inégalité $\Lambda(\mathbf{y}_n)\geqslant C$ (voir (3.53)).

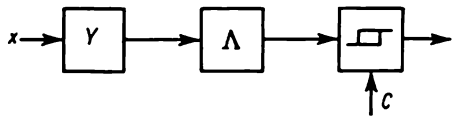



Fig. 3.2. Schéma fonctionnel du détecteur asymptotiquement optimal d'un signal quasi déterministe (observations indépendantes).

Dans le cas où l'on connaît les paramètres $\boldsymbol{\vartheta}'=(\vartheta_1,\ldots,\vartheta_m)$ (signal déterministe), la règle asymptotiquement optimale de détection prescrit en vertu de (3.49), (3.51) la comparaison avec un seuil de la statistique $\boldsymbol{\vartheta}'\mathbf{y}_n$ asymptotiquement normale de paramètres $(0,\ \mathrm{I}_f\boldsymbol{\vartheta}'\mathbf{A}\boldsymbol{\vartheta})$ pour l'hypothèse $H$ et de paramètres $(\gamma\mathrm{I}_f\boldsymbol{\vartheta}'\mathbf{A}\boldsymbol{\vartheta},\ \mathrm{I}_f\boldsymbol{\vartheta}'\mathbf{A}\boldsymbol{\vartheta})$ pour l'alternative $K$.

**3.1.7. Interprétation de l'algorithme asymptotiquement optimal de détection d'un signal quasi déterministe noyé dans un bruit indépendant.** En vertu de (3.53), le détecteur envisagé peut être réalisé à l'aide de deux blocs (fig. 3.2). Le premier bloc est un calculateur des composantes de la statistique vectorielle $\mathbf{y}_n$ d'après (3.51)

$$y_n^{(k)}(\mathbf{x})=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_k(t_i)\,f[x(t_i)]. \tag{3.57}$$

Le second bloc du détecteur sert à calculer, d'après une distribution donnée des paramètres du signal, le rapport de vraisemblance moyenne conformément à (3.52).

En comparant (3.57) et (3.9) on voit que le schéma de la figure 3.1 (sans dispositif à seuil) peut être utilisé en tant que bloc de calcul de la composante $k$ de la statistique $\mathbf{y}_n$, à condition de remplacer le signal déterministe $s(t)$ par le signal $\varphi_k(t)$. Ainsi, le premier bloc du détecteur d'un signal quasi déterministe dépendant de $m$

paramètres se compose de $m$ sous-blocs identiques (fig. 3.1, sans dispositif à seuil).

Notons que la caractéristique $f(x)$ des éléments non linéaires du premier bloc, ainsi que la grandeur du seuil dépendent de la forme du bruit et de son mode d'interaction avec le signal. La caractéristique du bloc de moyennage ne dépend que de la distribution des paramètres du signal.

**3.1.8. Algorithme asymptotiquement optimal de détection d'un signal modulé de phase aléatoire.** Considérons un signal à bande étroite $\lambda s(t)$ de phase initiale aléatoire $\varphi_0$, modulé en amplitude, en phase et en fréquence:

$$s(t) = a(t)\cos[\omega_0 t + \psi(t) - \varphi_0]. \tag{3.58}$$

Avec les désignations

$$\vartheta_1 = \cos\varphi_0,\ \vartheta_2 = \sin\varphi_0, \tag{3.59}$$

$$\varphi_1(t) = a(t)\cos[\omega_0 t + \psi(t)], \tag{3.60}$$

$$\varphi_2(t) = a(t)\sin[\omega_0 t + \psi(t)], \tag{3.60'}$$

e signal s'écrit comme suit:

$$s(t) = \vartheta_1\varphi_1(t) + \vartheta_2\varphi_2(t). \tag{3.61}$$

Il est clair que c'est un cas particulier d'un signal quasi déterministe (3.47) du type général pour $m = 2$.

Ainsi, dans le cas envisagé la statistique vectorielle $\mathbf{y}_n$ est bidimensionnelle de composantes

$$y_n^{(1)}(\mathrm{x}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} a(t_i) f[x(t_i)]\cos[\omega_0 t_i + \psi(t_i)], \tag{3.62}$$

$$y_n^{(2)}(\mathrm{x}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} a(t_i) f[x(t_i)]\sin[\omega_0 t_i + \psi(t_i)]. \tag{3.62'}$$

La distribution simultanée des variables aléatoires $y_n^{(1)}$, $y_n^{(2)}$ est asymptotiquement normale de matrice de corrélation $\mathrm{I}_f\mathrm{A}$, où $\mathrm{I}_f$ est donné par (3.5) et les éléments de la matrice A sont (voir (3.50))

$$a_{11} = \lim_{T\to\infty}\frac{1}{T}\int_0^T a^2(t)\cos^2[\omega_0(t) + \psi(t)]\,dt,$$

$$a_{12} = a_{21} = \lim_{T\to\infty}\frac{1}{T}\int_0^T a^2(t)\cos[\omega_0 t + \psi(t)]\sin[\omega_0 t + \psi(t)]\,dt,$$

$$a_{22} = \lim_{T\to\infty}\frac{1}{T}\int_0^T a^2(t)\sin^2[\omega_0 t + \psi(t)]\,dt.$$

Mais le spectre de la modulation d'amplitude étant étroit par rapport à la fréquence porteuse (ce qui permet de dire que le signal est à bande étroite), on a

$$a_{11} = a_{22} = \lim_{T\to\infty} \frac{1}{2T} \int_0^T a^2(t)\,dt = W_a/2, \tag{3.63}$$

$$a_{12} = a_{21} = 0, \tag{3.63'}$$

où $W_a$ est la puissance du processus de modulation en amplitude. La condition (3.63') signifie que les variables $y_n^{(1)}$ et $y_n^{(2)}$ sont indépendantes.

Les valeurs moyennes des variables $y_n^{(1)}$, $y_n^{(2)}$ sont nulles pour l'hypothèse $H$, et pour l'alternative $K$ sont égales à

$$m_1\{y_n^{(1)} \mid K\} = \frac{\gamma}{2} I_f W_a \cos\varphi_0, \tag{3.64}$$

$$m_1\{y_n^{(2)} \mid K\} = \frac{\gamma}{2} I_f W_a \sin\varphi_0. \tag{3.64'}$$

Supposons que la phase aléatoire $\varphi_0$ soit répartie uniformément sur l'intervalle $(0,2\pi)$. La distribution simultanée des variables $\vartheta_1 = \cos\varphi_0$. $\vartheta_2 = \sin\varphi_0$ est alors (voir tome I, page 95)

$$w_2(\vartheta_1, \vartheta_2) = \frac{1}{\pi\sqrt{1-\vartheta_1^2}}\,\delta\left(\vartheta_2 - \sqrt{1-\vartheta_1^2}\right). \tag{3.65}$$

Ensuite, dans le cas envisagé, on a

$$\vartheta' A \vartheta = \frac{W_a}{2}\cos^2\varphi_0 + \frac{W_a}{2}\sin^2\varphi_0 + \frac{W_a}{2}. \tag{3.66}$$

En substituant (3.62), (3.62'), (3.65), (3.66) dans (3.52) on obtient l'expression suivante pour la moyenne du rapport de vraisemblance :

$$\Lambda(\mathbf{y}_n) = \int_{-1}^{1}\int_{-1}^{1} \frac{1}{\pi\sqrt{1-\vartheta_1^2}}\,\delta\left(\vartheta_2 - \sqrt{1-\vartheta_1^2}\right) \times$$

$$\times \exp\left(\gamma\vartheta_1 y_n^{(1)} + \gamma\vartheta_2 y_n^{(2)} - \frac{\gamma^2}{4} I_f W_a\right) d\vartheta_1\, d\vartheta_2 =$$

$$= e^{-\gamma^2 I_f W_a/4} \frac{1}{2\pi} \int_0^{2\pi} \exp\left(\gamma y_n^{(1)}\cos\varphi + \gamma y_n^{(2)}\sin\varphi\right) d\varphi$$

ou

$$\Lambda(\mathbf{y}_n) = e^{-\gamma^2 I_f W_a/4}\, I_0\left(\gamma\sqrt{y_n^{(1)2} + y_n^{(2)2}}\right). \tag{3.67}$$

En vertu de (3.67), le rapport de vraisemblance moyenné est une fonction monotone de la statistique

$$|y_n|^2 = \frac{1}{n}\left[\left\{\sum_{i=1}^{n} a(t_i) f[x(t_i)] \cos[\omega_0 t_i + \psi(t_i)]\right\}^2 + \right.$$
$$\left. + \left\{\sum_{i=1}^{n} a(t_i) f[x(t_i)] \sin[\omega_0 t_i + \psi(t_i)]\right\}^2\right]. \qquad (3.68)$$

La formule (3.68) peut s'écrire sous la forme d'une somme double

$$|y_n|^2 = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} a(t_i)\, a(t_j)\, f[x(t_i)] \times$$
$$\times f[x(t_j)] \cos[\omega_0 (t_i - t_j) + \psi(t_i) - \psi(t_j)]. \qquad (3.69)$$

La règle asymptotiquement optimale de détection du signal (3.58) de phase aléatoire uniformément répartie s'énonce comme suit: on prend la décision que le signal est présent si *)

$$|y_n|^2 \geqslant C, \qquad (3.70)$$

et la décision que le signal est absent si c'est l'inégalité inverse de (3.70) qui est vérifiée. La statistique $|y_n|^2$ se calcule à l'aide de (3.68) ou (3.69). Notons qu'elle ne dépend pas de la constante $\gamma$.

Les variables asymptotiquement normales $y_n^{(1)}$, $y_n^{(2)}$ étant indépendantes, de moyennes nulles pour l'hypothèse $H$, la somme normée de leurs carrés est, pour cette hypothèse, répartie suivant la loi du $\chi^2$ à deux degrés de liberté (suivant la loi exponentielle). Autrement dit, la densité de probabilité de la variable aléatoire $2\,|y_n|^2/(I_f W_a)$ est égale à

$$w(z) = 1/2 e^{-z/2}, \quad z > 0. \qquad (3.71)$$

Ceci permet de trouver très simplement la grandeur du seuil $C$ dans l'inégalité (3.70), car pour une probabilité donnée de fausse alarme $\alpha$ on a

$$P\left\{\frac{2\,|y_n|^2}{I_f W_a} \geqslant \frac{2C}{I_f W_a}\,\middle|\, H\right\} = 1/2 \int_{2C/(I_f W_a)}^{\infty} e^{-z/2}\, dz = \exp\left\{\frac{-2C}{I_f W_a}\right\} = \alpha,$$

d'où l'on trouve

$$C = (I_f W_a/2) \ln 1/\alpha. \qquad (3.72)$$

*) Si dans (3.69) on rejette les composantes haute fréquence, (3.70) peut s'écrire comme suit:

$$\frac{1}{n}\sum_{i=1}^{n} a^2(t_i) f^2[x(t_i)] \sim \frac{1}{T}\int_0^T a^2(t) f^2[x(t)]\, dt \geqslant C.$$

Il est assez difficile de calculer la probabilité de détection correcte, car pour l'alternative $K$ les moyennes des variables aléatoires asymptotiquement normales $y_n^{(1)}$ et $y_n^{(2)}$ qui restent indépendantes, ne sont plus nulles comme il ressort de (3.64) et (3.64'). Dans ce cas, la densité de probabilité de la somme normée des carrés de $y_n^{(1)}$ et $y_n^{(2)}$, c'est-à-dire de la grandeur $2\,|\,y_n\,|^2/(\mathrm{I}_f\mathrm{W}_a)$ est de la forme (voir tome I, page 356):

$$w(z)=\frac{1}{2}\exp\left[-\left(z+\frac{\gamma^2}{2}\mathrm{I}_f\mathrm{W}_a\right)\Big/2\right]\mathrm{I}_0\times$$
$$\times\left(\gamma\sqrt{\frac{\mathrm{I}_f\mathrm{W}_a z}{2}}\right),\quad z>0. \qquad (3.73)$$

En posant $y^2 = z$, la fonction (3.73) devient la fonction généralisée de la repartition de Rayleigh (voir tome I, page 107). La probabilité de détection correcte est

$$1-\beta=\frac{1}{2}\int\limits_{2C/(\mathrm{I}_f\mathrm{W}_a)}^{\infty}\exp\left[-\left(z+\frac{\gamma^2}{2}\mathrm{I}_f\mathrm{W}_a\right)\Big/2\right]I_0\left(\gamma\sqrt{\mathrm{I}_f\mathrm{W}_a z/2}\right)dz,$$

d'où

$$z_{1-\beta}=\sqrt{\frac{2C}{\mathrm{I}_f\mathrm{W}_a}}=\sqrt{\ln\frac{1}{\alpha}}, \qquad (3.74)$$

où $z_{1-\beta}$ est l'1 — $\beta$-quantile de la loi généralisée normée de Rayleigh de paramètre $\gamma^2\mathrm{I}_f\mathrm{W}_a/2$.

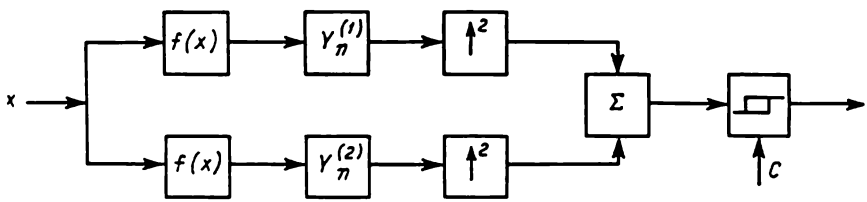



Fig. 3.3. Schéma fonctionnel du détecteur asymptotiquement optimal d'un signal à bande étroite (observations indépendantes).

En vertu de (3.68), le détecteur asymptotiquement optimal d'un signal modulé de phase aléatoire uniformément répartie doit avoir deux canaux en quadrature, où sont calculées les sommes corrélationnelles de la forme (3.9), l'une suivant les composantes en phase et l'autre des composantes en quadrature du signal. Puis, comme toujours pour la réception incohérente, on a des détecteurs quadratiques, un sommateur et un dispositif à seuil (fig. 3.3). Le schéma diffère du schéma analogue d'un détecteur d'un bruit normal additif par la présence de convertisseurs non linéaires d'entrée de caracté-

ristique $f(x)$. Pour conclure, remarquons qu'il est possible de ne pas exiger que la phase aléatoire $\varphi_0$ du signal soit uniformément répartie. Il suffit de supposer que la phase $\varphi_0$ soit une somme de phases indépendantes, l'une ayant une répartition quelconque et l'autre suivant une loi uniforme. Dans ce cas, on peut utiliser la proposition suivante (voir problème 3.7 dans le tome I): soit $\xi$ la somme de deux variables aléatoires indépendantes continues dont l'une est uniformément répartie sur l'intervalle $(-a, a)$ et l'autre ayant une répartition arbitraire. La densité de probabilité de $\xi$, rapportée à l'intervalle $(-a, a)$, est alors *toujours uniforme*.

**3.1.9. Algorithme postdétecteur asymptotiquement optimal de détection d'un signal modulé de phase aléatoire (méthode d'amplitude).** Nous allons passer à l'étude des algorithmes de détection de signaux appliqués à des processus à bande étroite ayant subi, avant le traitement optimal, une démodulation d'amplitude ou de phase. Considérons tout d'abord l'algorithme asymptotiquement optimal de détection du type (3.58) utilisant la statistique des valeurs échantillonnées indépendantes qui sont des valeurs de l'enveloppe du processus observé.

Soit $\mathbf{r} = (r_1, \ldots, r_n)$ un échantillon de l'enveloppe du processus observé qui est soit l'enveloppe d'un bruit stationnaire aléatoire à bande étroite, soit l'enveloppe du mélange du bruit et du signal (3.58). Désignons par $W_1(r; 0)$ la fonction de répartition de l'enveloppe du bruit et par $W_1(r; \vartheta)$ la fonction de répartition de l'enveloppe du mélange du signal (3.58) et du bruit avec

$$\vartheta = \lambda_a a^2(t) = \gamma_a a^2(t)/\sqrt{n}, \ \lambda_a > 0, \tag{3.75}$$

pour un instant $t$ donné, où $\gamma_a = \lambda_a \sqrt{n}$ est une constante positive bornée. En vertu de (3.75) le paramètre $\vartheta$ est proportionnel au carré de l'enveloppe du signal. Comme la fonction de répartition du mélange de l'enveloppe et du bruit dépend du paramètre non négatif $\vartheta$, la forme (3.75) est naturelle parce que le développement de cette fonction en série de Taylor suivant les puissances de $a$ ne contient pas de puissances impaires. C'est pourquoi en adoptant (3.75) au lieu de $\vartheta = \lambda a(t)$ on évite que la dérivée $\frac{\partial}{\partial\vartheta} W_1(r; \vartheta)$ soit identiquement nulle pour $\vartheta = 0$ (voir (3.4)).

On peut maintenant, en utilisant les résultats de 3.1.3, écrire l'algorithme asymptotiquement optimal postdétecteur (cas de la démodulation d'amplitude) du signal (3.58) noyé dans un bruit stationnaire

$$\Phi(\mathbf{r}) = 1 \text{ (le signal est présent) lorsque } y_n(\mathbf{r}) \geqslant C,$$
$$\Phi(\mathbf{r}) = 0 \text{ (le signal est absent) lorsque } y_n(\mathbf{r}) < C, \tag{3.76}$$

où

$$y_n(\mathbf{r}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} a_i^2 f_a(r_i); \tag{3.77}$$

$$a_i = a(t_i), \quad i = 1, \ldots, n. \tag{3.78}$$

$$f_a(r) = \frac{1}{W_1(r;0)} \frac{\partial}{\partial \vartheta} W_1(r;\vartheta)|_{\vartheta=0}, \quad f_a(r) \not\equiv 0, \tag{3.79}$$

avec

$$m_1\{f_a(r) \mid H\} = \int_0^\infty f_a(r) W_1(r;0)\, dr = 0, \tag{3.80}$$

$$\mathrm{I}_{f_a} = m_1\{f_a^2(r) \mid H\} = \int_0^\infty f_a^2(r) W_1(r;0)\, dr < 0. \tag{3.81}$$

La statistique (3.77) est asymptotiquement normale de paramètres $(0, \mathrm{W}_{a^2}\mathrm{I}_{f_a})$ pour l'hypothèse $H$ et de paramètres $(\gamma_a \mathrm{W}_{a^2}\mathrm{I}_{f_a}, \mathrm{W}_{a^2}\mathrm{I}_{f_a})$ pour l'alternative $K$, où

$$\mathrm{W}_{a^2} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} a_i^4 \sim \lim_{T\to\infty} \frac{1}{T} \int_0^\infty a^4(t)\, dt, \quad \mathrm{W}_{a^2} < 0. \tag{3.82}$$

Le seuil $C$ dans (3.76) pour une probabilité donnée de fausse alarme $\alpha$ est donné par

$$C = x_\alpha \sqrt{\mathrm{I}_{f_a}\mathrm{W}_{a^2}}, \tag{3.82'}$$

où $x_\alpha$ est l'$\alpha$-quantile de la loi normale.

La probabilité $\beta$ de perte du signal est liée à la probabilité $\alpha$ de fausse alarme par une relation analogue à (3.18):

$$(x_\alpha - x_{1-\beta})^2 = \gamma_a^2 \mathrm{I}_{f_a} \mathrm{W}_{a^2}^2. \tag{3.83}$$

Le coefficient d'efficacité asymptotique relative de l'algorithme postdétecteur (3.77) par rapport à l'algorithme prédétecteur (3.13) est, en vertu de (3.18') et (3.83), égal à

$$\rho = \frac{\mathrm{W}_{a^2}\mathrm{I}_{f_a}}{\mathrm{W}_a \mathrm{I}_f} \lim_{\lambda\to 0} \frac{\lambda_a^2}{\lambda^2} = 0$$

car $\lambda_a^2 = \lambda^4$.

Envisageons à titre d'exemple le problème de la détection du signal (3.58) noyé dans un bruit normal additif à bande étroite de valeur moyenne nulle et de variance $\sigma^2$. Dans ce cas, dans l'enveloppe, le signal et le bruit ne sont pas additifs. L'enveloppe du mélange du signal et du bruit est répartie suivant la loi généralisée de Rayleigh

$$W_1(r;\vartheta) = \frac{r}{\sigma^2} e^{-\frac{r^2+\vartheta}{2\sigma^2}} \mathrm{I}_0\left(\frac{r\sqrt{\vartheta}}{\sigma^2}\right), \quad r \geqslant 0, \tag{3.84}$$

et l'enveloppe du bruit suit la loi de Rayleigh

$$W_1(r;0)=\frac{r}{\sigma^2}e^{-r^2/2\sigma^2},\quad r>0. \tag{3.85}$$

Portons (3.84), (3.85) dans (3.79), il vient

$$f_a(r)=\frac{\partial}{\partial\vartheta}\left[e^{-\vartheta/2\sigma^2}\mathrm{I}_0\left(\frac{r\sqrt{\vartheta}}{\sigma^2}\right)\right]_{\vartheta=0}=$$

$$=\left[-\frac{1}{2\sigma^2}e^{-\vartheta/2\sigma^2}\mathrm{I}_0\left(\frac{r\sqrt{\vartheta}}{\sigma^2}\right)+e^{-\frac{\vartheta^2}{2\sigma^2}}\frac{r}{2\sigma^2\sqrt{\vartheta}}\mathrm{I}_1\left(\frac{r\sqrt{\vartheta}}{\sigma^2}\right)\right]_{\vartheta=0}$$

et compte tenu de $\mathrm{I}_1(x)=\frac{x}{2}+o(x)$ pour des $x$ petits, on obtient

$$f_a(r)=(r^4-2\sigma^2)/(4\sigma^6). \tag{3.86}$$

En substituant (3.86) dans (3.77), on obtient la règle asymptotiquement optimale suivante de détection du signal (3.58) dans un bruit additif stationnaire: le signal est présent si

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}a_i^2r_i^2\geqslant 4\sigma^6C+\frac{2\sigma^2}{\sqrt{n}}\sum_{i=1}^{n}a_i^2. \tag{3.87}$$

Dans ce cas la quantité d'information d'après Fisher est en vertu de (3.81) et (3.86) égale à

$$\mathrm{I}_{f_a}=\frac{1}{4\sigma^4}\int_0^\infty u\left(\frac{u^2}{2}-1\right)^2e^{-u^2/2}\,du=\frac{1}{4\sigma^4}; \tag{3.88}$$

à l'aide de (3.82') l'inégalité (3.87) peut alors s'écrire

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}a_i^2r_i^2\geqslant 2\sigma^2\left(x_\alpha\sqrt{\mathrm{W}_{a^2}}+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}a_i^2\right). \tag{3.89}$$

**3.1.10. Algorithme asymptotiquement optimal postdétecteur de détection d'un signal modulé de phase aléatoire (méthode de phase).** Considérons l'algorithme asymptotiquement optimal de détection d'un signal du type (3.58), utilisant une statistique de valeurs échantillonnées indépendantes $\theta=(\theta_1,\ldots,\theta_n)$ de la phase du processus observé, qui peut être soit la phase d'un bruit stationnaire à bande étroite, soit la phase du mélange de ce bruit et du signal.

Fixons tout d'abord la valeur de la phase aléatoire ($\varphi_0=0$) et désignons par $W_1(\theta;0)$ la fonction de probabilité de la phase du bruit, et par $W_1(\theta;\vartheta)$ celle de la phase du mélange du signal et du bruit et remarquons que pour un instant donné $t$ on a

$$\vartheta=\lambda a(t)=\gamma a(t)/\sqrt{n},\quad \lambda>0, \tag{3.90}$$

où $\gamma = \lambda \sqrt{n}$ est une constante positive bornée. A la différence de (3.75), dans le cas envisagé le paramètre $\vartheta$ est proportionnel aux valeurs de l'enveloppe car le développement $W_1(\theta; \lambda a)$ contient un terme linéaire en $a$.

En utilisant les résultats de 3.1.3, nous pouvons écrire l'algorithme postdétecteur (après démodulation de phase) asymptotiquement optimal de détection d'un signal du type (3.58) (pour $\varphi_0 = 0$) noyé dans un bruit stationnaire:

$$\Phi(\theta) = 1 \text{ (le signal est présent)}, \quad y_n(\theta) \geqslant C,$$
$$\Phi(\theta) = 0 \text{ (le signal est absent)}, \quad y_n(\theta) < C, \tag{3.91}$$

où

$$y_n(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} a_i f_\psi(\theta_i), \quad a_i = a(t_i), \tag{3.92}$$

$$f_\psi(\theta) = \frac{1}{W_1(\theta; 0)} \frac{\partial}{\partial \vartheta} W_1(\theta, \vartheta)\Big|_{\vartheta=0}, \quad f_\psi(\theta) \not\equiv 0, \tag{3.93}$$

avec

$$m_1\{f_\psi(\theta) \mid H\} = \int_{-\pi}^{\pi} f_\psi(\theta) W_1(\theta, 0)\, d\theta = 0, \tag{3.94}$$

$$\mathrm{I}_{f_\psi} = m_1\{f_\psi^2(\theta) \mid H\} = \int_{-\pi}^{\pi} f_\psi^2(\theta) W_1(\theta, 0)\, d\theta < \infty. \tag{3.95}$$

La statistique (3.92) est asymptotiquement normale de paramètres $(0, \mathrm{W}_a \mathrm{I}_{f_\psi}/2)$ pour l'hypothèse $H$, et de paramètres $(\gamma \mathrm{W}_a \mathrm{I}_{f_\psi}/2, \mathrm{W}_a \mathrm{I}_{f_\psi}/2)$ pour l'alternative $K$, où

$$\frac{\mathrm{W}_a}{2} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} a_i^2 \sim \lim_{T\to\infty} \frac{1}{2T} \int_0^T a^2(t)\, dt, \tag{3.96}$$

car la bande passante $F_a$ de l'enveloppe est deux fois plus étroite que la bande passante $F$ du signal ($n = 2FT = 4F_aT$).

La grandeur du seuil dans (3.91) pour une probabilité donnée $\alpha$ de fausse alarme est donnée par

$$C = x_\alpha \sqrt{\mathrm{I}_{f_\psi} \mathrm{W}_a/2}, \tag{3.97}$$

où $x_\alpha$ est l'$\alpha$-quantile de la loi normale. La probabilité $\beta$ de perte du signal est liée à la probabilité de fausse alarme $\alpha$ par une relation analogue à (3.83):

$$(x_\alpha - x_{1-\beta})^2 = \gamma^2 \mathrm{I}_{f_\psi} \mathrm{W}_a/2. \tag{3.98}$$

Le coefficient d'efficacité asymptotique relative de l'algorithme postdétecteur (3.92) (pour la phase initiale donnée) par rapport

à l'algorithme (3.13) est en vertu de (3.18) et (3.98) égal à

$$\rho = I_{f_\psi}/I_f, \tag{3.99}$$

car

$$W_s = \lim_{T\to\infty} \frac{1}{T} \int_0^T s^2(t)\,dt = \lim_{T\to\infty} \int_0^T a^2(t) \cos^2 \omega_0 t\,dt = \frac{1}{2} W_a.$$

Si la phase $\varphi_0$ est aléatoire et uniformément répartie sur l'intervalle $(-\pi, \pi)$, en vertu de (3.52), en prenant la moyenne sur la phase, on obtient:

$$\Lambda[y_n(\theta)] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp\left\{\gamma y_n(\theta - \varphi_0) - \frac{\gamma^2}{2} I_{f_\psi} W_a\right\} d\varphi_0. \tag{3.100}$$

La règle asymptotiquement optimale de détection après la démodulation de phase se formule comme suit: le signal est présent si

$$\Lambda[y_n(\theta)] \geqslant C. \tag{3.101}$$

La valeur du seuil $C$ et la constante $\gamma$ satisfont au système d'équations

$$\begin{aligned} P\{\Lambda[y_n(\theta)] \geqslant C \mid H\} &= \alpha, \\ P\{\Lambda[y_n(\theta)] < C \mid K\} &= \beta. \end{aligned} \tag{3.102}$$

A titre d'exemple considérons le cas d'un bruit normal additif à bande étroite de valeur moyenne nulle et de variance $\sigma^2$. La distribution de la phase du mélange du signal (3.58) (pour $\varphi_0 = 0$) et du bruit est dans ce cas (voir tome I, page 362)

$$\begin{aligned} W_1(\theta, \vartheta) = \frac{1}{2\pi} e^{-\vartheta^2/2\sigma^2} + \\ + \frac{\vartheta}{\sigma\sqrt{2\pi}} \cos(\theta - \psi) F\left[\frac{\vartheta}{\sigma} \cos(\theta - \psi)\right] \exp\left\{-\frac{\vartheta^2}{2\sigma^2} \sin^2(\theta - \psi)\right\}, \\ |\vartheta - \psi| \leqslant \pi, \end{aligned} \tag{3.103}$$

la distribution de la phase du bruit étant uniforme sur l'intervalle $(-\pi, \pi)$:

$$W_1(\theta; 0) = \frac{1}{2\pi}. \tag{3.103'}$$

A partir de (3.93) on obtient

$$f_\psi(\theta) = \frac{1}{\sigma} \sqrt{\frac{\pi}{2}} \cos(\theta - \psi) \tag{3.104}$$

et par conséquent *)

$$y_n(\theta) = \sqrt{\frac{\pi}{2n}} \sum_{i=1}^{n} \frac{a_i}{\sigma} \cos(\theta_i - \psi_i) = \sqrt{\frac{\pi}{2}} A_n \cos \Phi_n, \quad (3.105)$$

où

$$A_n = \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{a_j}{\sigma} e^{i(\theta_j - \psi_j)} \right|; \quad (3.106)$$

$$\Phi_n = \operatorname{arc\,tg} \left[ \operatorname{Im} \sum_{j=1}^{n} a_j e^{i(\theta_j - \psi_j)} / \operatorname{Re} \sum_{j=1}^{n} a_j e^{i(\theta_j - \psi_j)} \right]. \quad (3.107)$$

La quantité d'information d'après Fisher est

$$I_{f_\psi} = \frac{1}{4\sigma^2} \int_{-\pi}^{\pi} \cos^2(\theta - \psi)\, d\theta = \frac{\pi}{4\sigma^2}. \quad (3.108)$$

Dans ce cas (pour une phase initiale donnée) le coefficient d'efficacité asymptotique relative de l'algorithme postdétecteur (3.105) par rapport à l'algorithme prédétecteur (linéaire pour un bruit normal additif) conformément à (3.99) et (3.108) est égal à $\frac{\pi}{4} \approx 0{,}78$.

Si la phase $\varphi_0$ du signal est aléatoire, en exprimant $y_n(\theta - \varphi_0)$ de (3.100) par (3.105), portant (3.108) dans (3.100) et en prenant la moyenne sur la phase uniforme, on obtient:

$$\Lambda[y_n(\theta)] = I_0\left(\gamma \sqrt{\frac{\pi}{2}} A_n\right) e^{-\frac{\pi^2}{8}\gamma^2 W_a/\sigma^2}. \quad (3.109)$$

Le rapport de vraisemblance moyenné $\Lambda[y_n(\theta)]$ étant une fonction monotone de la statistique $A_n$, l'algorithme postdétecteur (de phase) asymptotiquement optimal peut s'écrire comme suit **): le signal est présent si

$$\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{a_j}{\sigma} e^{i(\theta_j - \psi_j)} \right| \geqslant C. \quad (3.110)$$

La distribution de la variable aléatoire $A_n$ tend vers une loi de Rayleigh de paramètre $W_a/(2\sigma^2)$ en l'absence de signal (hypo-

*) Notons que l'algorithme (3.105) est non paramétrique par rapport à la classe des bruits caractérisés par une densité de probabilité simultanée de l'enveloppe et de la phase de la forme $w_2(r, \theta) = \frac{1}{2\pi} w_1(r)$ pour $w_1(r)$ quelconque (voir [17]).

**) Remarquons que l'algorithme (3.110) est non paramétrique dans le même sens que l'algorithme basé sur la statistique (3.105).

thèse $H$) et vers une loi de Rayleigh généralisée de paramètres $\gamma\sqrt{\frac{\pi W_a}{8\sigma^2}}$, $\frac{W_a}{2\sigma^2}$ en présence du signal (alternative $K$) (voir tome II, page 304). Par conséquent, pour des probabilités données $\alpha$ de fausse alarme et $\beta$ de perte du signal, les grandeurs $C$ et $\gamma$ satisfont aux équations suivantes:

$$C=\frac{W_a}{2\sigma^2}\ln\frac{1}{\alpha}, \tag{3.111}$$

$$\beta=\exp\left(-\frac{\pi}{8}\gamma^2\frac{W_a}{\sigma^2}\right)\int\limits_0^{\sqrt{\ln 1/\alpha}} xe^{-x^2/2}I_0\left(\gamma\sqrt{\frac{\pi}{8}\frac{W_a}{\sigma}}\,x\right)dx. \tag{3.112}$$

**3.1.11. Distribution normale des paramètres du signal.** Revenons aux algorithmes prédétecteurs asymptotiquement optimaux. Considérons maintenant le cas des signaux avec un nombre quelconque de paramètres indépendants et supposons que leur distribution simultanée soit normale, de vecteurs des moyennes $\vartheta_0$ et de matrice de corrélation diagonale $DI$ donnés. En utilisant (1.33) avec des changements de désignations requis, on trouve que dans le cas envisagé le rapport de vraisemblance moyenné (3.52) est une fonction monotone de la statistique

$$\psi_1(y_n)=y_n'A^{-1}y_n-(y_n-\gamma I_f A\vartheta_0)'\left(A+\gamma^2\frac{I_f}{D}A'A\right)^{-1}(y_n-\gamma I_f A\vartheta_0). \tag{3.113}$$

Les éléments de la matrice A sont donnés par (3.50) et la statistique $y_n$ par (3.51).

Avant de formuler la règle de détection du signal et de chercher la distribution de la statistique $\psi_1(y_n)$ qui nous permettra de trouver le seuil $C$ et la constante $\gamma$, nous allons restreindre la classe des fonctions de base $\varphi_i(t)$ du signal. Supposons que ces fonctions soient asymptotiquement orthogonales et normées, c'est-à-dire

$$\lim_{T\to\infty}\frac{1}{T}\int\limits_0^T \varphi_j(t)\,\varphi_k(t)\,dt=\delta_{jk}, \tag{3.114}$$

où $\delta_{jk}$ est le symbole de Kronecker. En vertu de (3.114), la matrice A est unitaire. La statistique (3.113) s'écrit alors

$$\psi_1(y_n)=y_n'y_n-\frac{D}{D+\gamma^2 I_f}(y_n-\gamma I_f\vartheta_0)'(y_n-\gamma I_f\vartheta_0). \tag{3.115}$$

En rejetant dans (3.115) les termes constants qui ne dépendent pas de la statistique $y_n$ et en complétant l'expression qui reste

jusqu'à obtenir un carré complet on arrive à

$$\psi_1(\mathbf{y}_n) = \left|\mathbf{y}_n + \frac{D}{\gamma}\vartheta_0\right|^2 = \sum_{j=1}^{m}\left(y_n^{(j)} + \frac{D}{\gamma}\vartheta_{0j}\right)^2, \qquad (3.116)$$

où $y_n^{(i)}$ est la composante de la statistique $\mathbf{y}_n$ correspondant à la $j$-ième fonction de l'échantillon du signal (voir (3.57)); $\vartheta_{0j}$ est la $j$-ième composante du vecteur des valeurs moyennes des paramètres du signal.

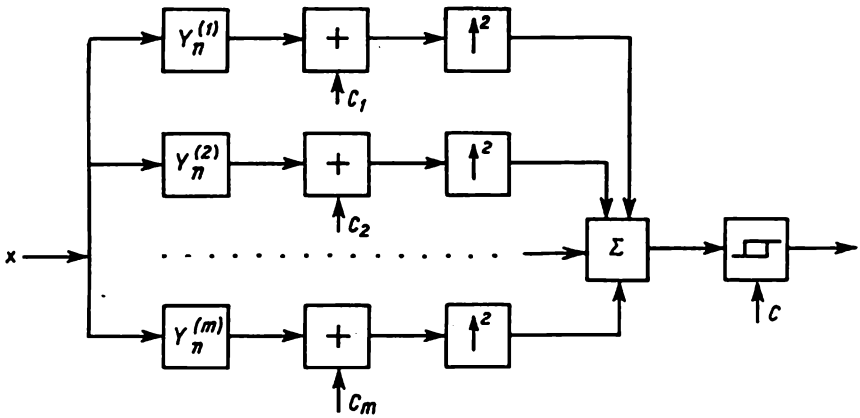



Fig. 3.4. Schéma fonctionnel du détecteur asymptotiquement optimal d'un signal dans un processus ayant subi une démodulation préalable (observations indépendantes).

La règle asymptotiquement optimale de détection d'un signal peut maintenant être formulée comme suit: le signal est présent si

$$\frac{1}{I_j}\sum_{j=1}^{m}\left(y_n^{(j)} + \frac{D}{\gamma}\vartheta_{0j}\right)^2 \geqslant C, \qquad (3.117)$$

et il n'y a pas de signal si c'est l'inégalité inverse qui est vérifiée.

La statistique (3.117) est la somme des carrés de variables aléatoires indépendantes asymptotiquement normales de même variance (unité) et de valeurs moyennes différentes. Pour l'hypothèse $H$, lorsque les valeurs moyennes des $y_n^{(j)}$ sont nulles, cette statistique est répartie suivant une loi du $\chi^2$ non centrée à $m$ degrés de liberté [22], le carré du paramètre d'écart étant égal à

$$\delta_H^2(\gamma) = \frac{D^2}{\gamma^2 I_j}\,|\vartheta_0|^2. \qquad (3.118)$$

La réalisation pratique de l'algorithme (3.117) est assez simple (fig. 3.4). Les $m$ calculateurs spéciaux fournissent les composantes

du vecteur $\mathbf{y}_n$ auxquelles s'ajoutent des constantes données, les sommes attaquent des dispositifs d'élévation au carré puis un sommateur commun. La valeur de la somme finale est comparée à un seuil.

La densité de probabilité $W_1(z;\ m,\ \delta)$ de la loi du $\chi^2$ non centrée peut être exprimée en fonction de la densité de probabilité $W_1(z;\ l,\ 0)$ de la loi du $\chi^2$ centrée:

$$W_1(z;\ m,\ \delta) = e^{-\delta^2/2} \sum_{l=0}^{\infty} \frac{1}{l!} \left(\frac{\delta^2}{2}\right)^l W_1(z;\ m+2l,\ 0). \qquad (3.119)$$

Pour $\vartheta$ donné et l'alternative $K$, la valeur moyenne du vecteur $\mathbf{y}_n$ est égale à $\gamma \mathrm{I}_f \vartheta$ et par conséquent la statistique (3.117) est également répartie suivant une loi du $\chi^2$ non centrée à $m$ degrés de liberté, de paramètre d'écart égal à

$$\delta_K^2(\gamma) = \sum_{i=1}^{m} \left(\gamma \sqrt{\mathrm{I}_f}\, \vartheta_i + \frac{D}{\gamma \sqrt{\mathrm{I}_f}}\, \vartheta_{0i}\right)^2. \qquad (3.120)$$

Si $\vartheta_i$ sont des variables aléatoires indépendantes normales de même variance $D$ et de moyennes $\vartheta_{0i}$, $i = 1, \ldots, m$, alors $\delta_K^2(\gamma)/(\gamma^2 \mathrm{I}_f D)$ est une variable aléatoire qui suit une loi du $\chi^2$ non centrée à $m$ degrés de liberté et de paramètre d'écart égal à

$$\delta_K^{*2}(\gamma) = \frac{1}{D}\left(1 + \frac{D}{\gamma^2 \mathrm{I}_f}\right)^2 |\vartheta_0|^2. \qquad (3.121)$$

Le moyennage de la densité de la loi du $\chi^2$ non centrée suivant le paramètre d'écart donne la double somme suivante contenant la densité de la loi du $\chi^2$ centrée (comparer avec (3.119)):

$$W_1(z;\ m,\ \delta_K^*) = \int_0^{\infty} W_1(z;\ m,\ \delta_K)\, W_1(\delta_K)\, d\delta_K =$$

$$= \sum_{r=0}^{m} \sum_{l=0}^{m} \frac{2^{-\left(\frac{m+1}{2}+l+r\right)}}{\Gamma(r+1)\,\Gamma(l+1)\,\Gamma\left(\frac{m+2l}{2}\right)} \left(\frac{\delta_K^*}{2}\right)^l e^{-\delta_K^{*2}} W_1(z;\ m+2r,\ 0). \qquad (3.122)$$

On peut écrire maintenant le système d'équations donnant la constante $\gamma$ et le seuil $C$ dans l'algorithme (3.117) pour des probabilités données $\alpha$ de fausse alarme et $1 - \beta$ de détection correcte. On a

$$\alpha = P\{\chi^2(m,\ \delta) \geqslant C \mid H\} = \int_C^{\infty} W_1(z;\ m,\ \delta_H)\, dz, \qquad (3.123)$$

d'où l'on tire

$$C = \chi^2_\alpha [m, \delta^2_H(\gamma)], \tag{3.124}$$

c'est-à-dire que le seuil coïncide avec l'$\alpha$-quantile de la loi du $\chi^2$ non centrée dont le paramètre d'écart est donné par la formule (3.118). La probabilité de détection correcte se trouve à partir de (3.122):

$$1 - \beta = \int_0^\infty W_1(z;\ m,\ \delta^*_K)\, dz,$$

d'où

$$C = \varkappa_{1-\beta} [m, \delta^*_K(\gamma)], \tag{3.125}$$

où $\varkappa_{1-\beta}$ est l'1 — $\beta$-quantile de la loi donnée par (3.122). Tenant compte de (3.124), on peut écrire (3.125) comme suit:

$$\chi^2_\alpha [m, \delta_H(\gamma)] = \varkappa_{1-\beta} [m, \delta^{*2}_K(\gamma)] \tag{3.126}$$

et utiliser (3.126) pour calculer la constante $\gamma$ et puis le seuil $C$ conformément à (3.124).

**3.1.12. Détection d'un signal stochastique.** Les résultats du paragraphe précédent peuvent être utilisés pour trouver l'algorithme asymptotiquement optimal de détection, dans un bruit indépendant, d'un signal stochastique $\xi(t)$ qui représente un processus aléatoire normal de moyenne nulle et de fonction de corrélation donnée $B(t, y)$. On sait (voir tome II, § 3.3) que ce signal peut s'écrire comme un développement orthogonal sur l'intervalle d'observation $(0, T)$

$$\xi(t) = \sum_{i=1}^\infty \vartheta_i \varphi_i(t), \tag{3.127}$$

$$\vartheta_i = \int_0^T \xi(u)\, \varphi_i(u)\, du, \tag{3.127'}$$

où $\{\varphi_i(t)\}$ est la base orthonormée déterminée par les fonctions propres de l'équation intégrale homogène linéaire

$$\varphi(t) = \lambda \int_0^T B(t, y)\, \varphi(y)\, dy, \quad 0 \leqslant t \leqslant T. \tag{3.128}$$

Les paramètres $\vartheta_1, \vartheta_2, \ldots$, forment un ensemble (dénombrable) de variables aléatoires normales indépendantes de moyennes nulles

et de variances $1/\lambda_1$, $1/\lambda_2$, . . ., car en vertu de (3.128) on a

$$m_1\{\vartheta_i\vartheta_j\}=m_1\left\{\int_0^T\int_0^T \xi(u)\,\xi(\vartheta)\,\varphi_i(u)\,\varphi_j(v)\,du\,dv\right\}=$$

$$=\int_0^T\int_0^T B(u,\,v)\,\varphi_i(u)\,\varphi_j(v)\,du\,dv=\frac{1}{\lambda_i}\int_0^T \varphi_i(u)\,\varphi_j(u)\,du=\frac{1}{\lambda_i}\,\delta_{ij}.$$

En comparant (3.127) avec (3.47) on remarque que le signal stochastique $\xi(t)$ a la même structure que le signal quasi déterministe si l'on pose $m=\infty$. La différence est en ce que le vecteur de dimension infinie $\vartheta$ a un vecteur moyenne nul et une matrice de corrélation $\mathbf{K}_\vartheta$ diagonale dont les éléments $1/\lambda_i$ ne sont pas égaux entre eux.

Dans le cas envisagé le rapport de vraisemblance moyenné (3.52) est une fonction monotone de la statistique

$$\psi_2(\mathbf{y}_n)=\mathbf{y}_n'\mathbf{y}_n-\mathbf{y}_n'(\mathbf{I}+\gamma^2\mathrm{I}_f\mathbf{K}_\vartheta)^{-1}\,\mathbf{y}_n. \tag{3.129}$$

En utilisant la règle d'addition des matrices, la matrice $\mathbf{K}_\vartheta$ étant diagonale, on peut écrire (3.129) comme suit:

$$\psi_2(\mathbf{y}_n)=\sum_{k=1}^{\infty}\frac{\gamma^2\mathrm{I}_f}{\lambda_k+\gamma^2\mathrm{I}_f}\,(y_n^{(k)})^2, \tag{3.130}$$

où $y_n^{(k)}$ est la composante du vecteur $\mathbf{y}_n$ donnée par (3.57).

Introduisons la fonction

$$h(t,\,y)=\gamma^2\mathrm{I}_f\sum_{j=1}^{\infty}\frac{\varphi_j(t)\,\varphi_j(y)}{\lambda_j+\gamma^2\mathrm{I}_f}. \tag{3.131}$$

Il est facile de voir par simple substitution que la fonction $h(t,\,y)$ satisfait à l'équation intégrale suivante:

$$\int_0^T h(v,\,y)\,B(y,\,t)\,dy+\frac{1}{\gamma^3\mathrm{I}_f}\,h(t,\,v)=$$

$$=B(t,\,v),\quad 0\leqslant t\leqslant T,\quad 0\leqslant v\leqslant T. \tag{3.132}$$

En vertu de (3.57), (3.130) et (3.131), la statistique $\psi_2(\mathbf{y}_n)$ peut dans le cas envisagé s'écrire comme suit:

$$\psi_2(\mathbf{y}_n)=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}h(t_i,\,t_j)\,f[x(t_i)]\,f[x(t_j)]. \tag{3.133}$$

L'algorithme asymptotiquement optimal de détection d'un signal stochastique (3.127) prescrit la comparaison de la statistique (3.133)

avec un seuil. Autrement dit, le détecteur asymptotiquement optimal d'un signal aléatoire normal, de fonction de corrélation donnée, noyé dans un bruit stationnaire indépendant, de fonction de répartition quelconque, réalise le filtrage non linéaire des données échantillonnées qui ont subi une transformation non linéaire $f(x)$, et une comparaison à un seuil. La fonction de pondération du filtre non linéaire *) dépend de la fonction de corrélation du signal (voir (3.132)).

Si le bruit est additif et réparti suivant une loi normale, on a $f(x) = x/\sigma^2$, et la règle de détection que nous venons de formuler coïncide avec la règle bien connue de détection d'un signal normal noyé dans un bruit blanc normal (voir tome II, 5.3.3).

## 3.2. ALGORITHMES DE RANG ASYMPTOTIQUEMENT OPTIMAUX DE DÉTECTION DES SIGNAUX NOYÉS DANS DES BRUITS INDÉPENDANTS

**3.2.1. Equivalence asymptotique des statistiques de rang et des statistiques non ordonnées.** Dans le second chapitre nous avons envisagé les algorithmes de rang et de signe-rang de détection des signaux pour des tailles finies de l'échantillon. La synthèse des algorithmes de rang et de signe-rang, optimaux d'après le critère de Neyman-Pearson, de détection des signaux pour des tailles finies de l'échantillon présentant des difficultés mathématiques insurmontables, on choisit l'algorithme d'une manière purement heuristique. A titre d'exemple d'algorithmes proposés d'une manière heuristique on peut citer les algorithmes de Wilcoxon, de Van der Waerden, celui des médianes et autres envisagés dans le chapitre 2. On comprend donc l'intérêt des méthodes *asymptotiques* pour la synthèse optimale des dispositifs de détection utilisant des algorithmes de rang.

Comme nous l'avons déjà dit (voir 1.2.10), l'algorithme asymptotiquement le plus efficace de détection d'un signal noyé dans un bruit ne peut être déterminé d'une manière univoque. Il peut exister toute une classe d'algorithmes de ce genre équivalents par leurs propriétés asymptotiques. Tous les algorithmes de cette classe pour des échantillons de grandes tailles ont mêmes caractéristiques de détection des signaux faibles. Sous certaines conditions, cette classe comporte également des algorithmes non paramétriques de détection des signaux (de rang et de signe-rang). Autrement dit, avec certaines limitations imposées à la structure des signaux et des bruits, il existe des algorithmes de rang et de signe-rang asymptotiquement les plus efficaces qui, de par les caractéristiques de détection, sont équivalents (pour des échantillons de taille importante et des signaux

*) Il s'agit d'un filtre non linéaire du second ordre (voir tome II, page 258 )

faibles) aux algorithmes simples *), optimaux d'après le critère de Neyman-Pearson.

A première vue, cette assertion semble peu fondée, car lorsque l'on passe d'un échantillon intègre à ses rangs, une partie de l'information obtenue lors de l'observation se perd irrémédiablement. Et pourtant nous allons prouver, avec toute rigueur mathématique, qu'il existe des algorithmes de rang asymptotiquement optimaux aussi bons que les algorithmes simples. Notre certitude est basée sur les propositions fondamentales suivantes de la mathématique statistique.

Comme l'élément de l'échantillon $x_i$ coïncide avec la statistique d'ordre $x^{(R_i)}$, la densité de probabilité conditionnelle de l'échantillon $x_i$ pour un rang donné est égale à la densité de $x^{(R_i)}$. Pour un échantillon de taille $n$ issu de la distribution $w_1(x)$ on a à l'aide de (2.40')

$$w_1(x_i \mid R_i = R) = w_1[x^{(R_i)}] = \frac{n!}{(R-1)!\,(n-R)!} \times$$
$$\times [F_1(x_i)]^{R-1} [1 - F_1(x_i)]^{n-R} w_1(x_i), \quad (3.134)$$

où $F_1(x) = \int\limits_{-\infty}^{x} w_1(x)\,dx$,

Introduisons la variable aléatoire

$$u^{(R_i)} = F_1[x^{(R_i)}]. \quad (3.135)$$

Comme nous l'avons déjà remarqué dans 2.1.2, si $F_1(x)$ est la fonction de répartition de l'échantillon $x_i$, la variable aléatoire $u(R_i)$ définie selon (3.135) est répartie uniformément sur l'intervalle (0, 1). En introduisant le changement de variable (3.135) dans (3.134), on obtient la densité de probabilité de la statistique d'ordre à partir de la répartition uniforme

$$w_1[u^{(R)}] = \frac{n!}{(R-1)!\,(n-R)!} u^{R-1} (1-u)^{n-R}, \quad 0 < u < 1$$
$$1 \leqslant R \leqslant n. \quad (3.136)$$

Il est facile de trouver la valeur moyenne et la variance de $u^{(R)}$. Comme pour $m$ et $n$ entiers positifs quelconques on a

$$\int\limits_0^1 x^m (1-x)^n\,dx = \frac{m!\,n!}{(m+n+1)!},$$

*) Nous employons le terme « algorithme simple » pour désigner un algorithme qui traite des échantillons tels quels, non spécialement ordonnés, et nous appliquerons le terme « algorithme de rang » à un algorithme prévu pour des échantillons ordonnés.— *Note du rédacteur.*

en vertu de (3.136)

$$m_1\{u^{(R)}\}=\frac{R}{n+1}, \quad 1 \leqslant R \leqslant n, \tag{3.137}$$

$$M_2\{u^{(R)}\}=\frac{R(n-R+1)}{(n+1)^2(n+2)}, \quad 1 \leqslant R \leqslant n. \tag{3.138}$$

A partir de (3.135), (3.137) et (3.138), on trouve

$$m_1\{F_1(x_i) \mid R_i = R\} = R/(n+1) \tag{3.139}$$

et pour $n$ suffisamment grand

$$M_2\{F_1(x_i) \mid R_i = R\} \sim R/n^2. \tag{3.140}$$

Autrement dit, pour $n$ suffisamment grand, la variable aléatoire $F_1(x_i)$ s'écarte peu, en probabilité, de sa valeur moyenne égale à $R_i/(n+1)$, ce qui découle par exemple de l'inégalité de Tchébychev (voir tome I, page 71). Il en découle l'équivalence asymptotique de l'élément de l'échantillon $x_i$ et de la valeur convertie du rang $F_1^{-1}\left(\frac{R_i}{n+1}\right)$, si toutefois $F_1^{-1}(z)$ est la fonction inverse de la fonction de répartition de la loi à laquelle appartient l'échantillon. Donc les algorithmes de rang asymptotiquement optimaux de détection des signaux noyés dans des bruits indépendants peuvent être obtenus à partir des algorithmes asymptotiquement optimaux simples par la substitution

$$x_i \sim F_1^{-1}\left(\frac{R_i}{n+1}\right). \tag{3.141}$$

Nous allons maintenant formuler d'une manière plus rigoureuse les conditions pour lesquelles les statistiques de rang et simples sont asymptotiquement équivalentes. Définissons sur le segment (0, 1) la fonction $\Delta(u)$ satisfaisant aux conditions suivantes

$$\int_0^1 \Delta(u)\,du = 0, \quad \int_0^1 \Delta^2(u)\,du < \infty, \tag{3.142}$$

et la fonction en escalier

$$\Delta_n(u)=\Delta_n\left(\frac{k}{n+1}\right), \quad u \in \left(\frac{k-1}{n}, \frac{k}{n}\right); \quad k=1, \ldots, n, \tag{3.143}$$

avec

$$\lim_{n\to\infty} \int_0^1 [\Delta_n(u)-\Delta(u)]^2\,du = 0. \tag{3.144}$$

Soit $s_i$ une suite de nombres telle que, pour $n$ quelconque

$$\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i\right| < c < \infty, \tag{3.145}$$

$$\lim_{n\to\infty} \max_{1\leq i\leq n} s_i^2 / \sum_{i=1}^{n} s_i^2 = 0, \tag{3.145'}$$

et supposons que l'on ait un échantillon indépendant $u_1, \ldots, u_n$, issu d'une répartition normale sur l'intervalle (0, 1), $R_i$ étant le rang de l'élément $u_i$ dans cet échantillon. On peut alors montrer l'équivalence asymptotique des deux statistiques *) (voir par exemple [3])

$$\tilde{y}_n(u_1, \ldots, u_n) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i \Delta(u_i) \tag{3.146}$$

et

$$\tilde{z}_n(R_1, \ldots, R_n) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i \Delta_n\left(\frac{R_i}{n+1}\right), \tag{3.147}$$

c'est-à-dire

$$\tilde{y}_n(u_1, \ldots, u_n) = \tilde{z}_n(R_1, \ldots, R_n) + \chi_n, \tag{3.148}$$

où $\chi_n$ converge en probabilité vers zéro pour $n \to \infty$.

**3.2.2. Algorithme de rang asymptotiquement optimal de détection d'un signal déterministe.** Définissons la fonction $\Delta(u)$ introduite ci-dessus comme suit:

$$\Delta(u) = f[F_1^{-1}(u)], \quad 0 < u < 1, \tag{3.149}$$

où $(x)$ est une fonction donnée par la formule (3.4) et $F_1^{-1}(u)$ est la fonction inverse de la fonction de répartition du bruit

$$F_1(y) = \int_{-\infty}^{y} w_1(x;\ 0)\, dx. \tag{3.150}$$

Il est évident que la fonction $\Delta(u)$ définie par (3.149) satisfait aux conditions (3.142) car après le changement de variable $x = F^{-1}(u)$, on obtient

$$\int_0^1 f[F^{-1}(u)]\, du = \int_{-\infty}^{\infty} f(x)\, w_1(x \mid 0)\, dx = 0 \tag{3.151}$$

*) On prouve ici l'équivalence asymptotique des statistiques de rang et simple pour la répartition uniforme. Dans le paragraphe suivant nous allons montrer comment ce résultat peut être utilisé pour démontrer l'équivalence mentionnée pour une répartition quelconque.

et aussi (voir (3.5))

$$\int_0^1 f^2[F^{-1}(u)]\,du = \int_{-\infty}^{\infty} f^2(x)\,w_1(x;\,0)\,dx = \mathrm{I}_f < \infty. \qquad (3.152)$$

Comme nous l'avons déjà noté, à l'échantillon indépendant $x_1, \ldots, x_n$ du bruit de distribution $w_1(x;\,0)$ correspond l'échantillon indépendant $u_1, \ldots, u_n$ de la répartition uniforme si $x_i = F^{-1}(u_i)$. Soit $s_1, \ldots, s_n$ une suite de valeurs du signal déterministe $s(t)$, pour $t = t_1, \ldots, t_n$, satisfaisant aux conditions (3.145) et (3.145'). La première condition est vérifiée si le signal ne contient pas de composante constante, la seconde si les signaux sont réels. Dans ce cas la statistique (3.146) coïncide avec la statistique (3.9) utilisée pour la synthèse de l'algorithme asymptotiquement optimal de détection d'un signal déterministe noyé dans un bruit à valeurs indépendantes, car

$$\tilde{y}_n(u_1, \ldots, u_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i \Delta(u_i) =$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i f(x_i) = y_n(x_1, \ldots, x_n). \qquad (3.153)$$

L'équivalence asymptotique des statistiques (3.146) et (3.147) entraîne celle des statistiques

$$y_n(x_1, \ldots, x_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i f(x_i), \qquad (3.154)$$

$$z_n(R_1, \ldots, R_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i \Delta_n\left(\frac{R_i}{n+1}\right). \qquad (3.154')$$

Il ne reste plus qu'à utiliser la convergence en moyenne quadratique (voir (3.144)) pour $n \to \infty$

$$\Delta_n(u) \to \Delta(u) = f[F_1^{-1}(u)] \qquad (3.155)$$

pour établir l'équivalence asymptotique de la statistique $y_n(x_1, \ldots, x_n)$ déterminée par (3.154), et de la statistique de rang

$$z_n(R_1, \ldots, R_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i f\left[F_1^{-1}\left(\frac{R_i}{n+1}\right)\right]. \qquad (3.156)$$

L'équivalence asymptotique des statistiques $\tilde{y}_n$ et $z_n$ a lieu lorsque les $x_i$ ont même répartition de densité $w_1(x;\,0)$ donnée, c'est-à-dire en l'absence de signal. On peut cependant montrer que

cette assertion est également vraie dans le cas où l'échantillon appartient à la répartition

$$w_n(\mathbf{x};\ \gamma s/\sqrt{n}),\ 0<\gamma<\infty,\ \text{pour } n\to\infty.$$

En vertu du théorème de la limite centrale, la statistique de rang (3.156) est asymptotiquement normale tant pour l'hypothèse $H$ que pour l'alternative $K$ et tous les résultats formulés dans 3.1 pour la statistique (3.13) restent vrais pour (3.156).

L'algorithme de rang asymptotiquement optimal de détection d'un signal déterministe se formule maintenant comme suit: le signal est présent si

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i f\left[F_1^{-1}\left(\frac{R_i}{n+1}\right)\right]\geqslant C, \tag{3.157}$$

il n'y a pas de signal si c'est l'inégalité inverse qui est vérifiée.

Il est facile de voir que l'on peut obtenir (3.157) à partir de (3.13) en remplaçant $x_i$ par $F^{-1}\left(\frac{R_i}{n+1}\right)$ (voir (3.141)). Mais à la différence de (3.13), l'algorithme de détection basé sur (3.157) est un algorithme de rang, c'est-à-dire qu'il est *non paramétrique*, conservant le niveau de fausses alarmes pour un échantillon de taille quelconque, pour un seuil donné et une distribution du bruit variable.

L'algorithme de rang asymptotiquement optimal de détection d'un signal quasi déterministe du type général (3.47) (de composante constante nulle) peut être obtenu à partir de (3.52) en remplaçant la statistique $y_n(\mathbf{x})$ par la statistique $z_n(\mathbf{R})$ (voir (3.154) et (3.154')).

**3.2.3. Algorithmes de rang asymptotiquement optimaux de détection d'un signal déterministe noyé dans un bruit additif.** En vertu de (3.19), dans le cas d'un bruit additif, la caractéristique du convertisseur non linéaire dans (3.156) doit être de la forme

$$\Delta(u)=f[F_1^{-1}(u)]=w_1'[F_1^{-1}(u)]/w_1[F_1^{-1}(u)]. \tag{3.158}$$

Nous allons donner l'expression de $f[F_1^{-1}(u)]$ pour plusieurs distributions du bruit. Dans le cas d'un bruit normal, $f(x)=x/\sigma^2$ (voir (3.21)) et donc

$$f[F_1^{-1}(u)]=\frac{1}{\sigma^2}F_1^{-1}(u), \tag{3.159}$$

où $F_1^{-1}(u)$ est la fonction inverse de la fonction de répartition de la loi normale (intégrale de Laplace).

Pour un bruit laplacien on a $f(x) = \sqrt{\frac{2}{\sigma^2}}\,\mathrm{sgn}\,x$ (voir (3.23) et (3.24)) et donc

$$F_1(x) = \frac{1}{\sqrt{2\sigma^2}} \int_{-\infty}^{x} e^{-\sqrt{\frac{2}{\sigma^2}}|x|}\,dx =$$

$$= \frac{1}{2} \mp \frac{1}{2}\left(1 - e^{-\sqrt{\frac{2}{\sigma^2}}|x|}\right) = u, \qquad (3.160)$$

où le signe moins correspond à $x < 0$ et le signe plus à $x > 0$. A partir de (3.160) on trouve ensuite

$$e^{-\sqrt{\frac{2}{\sigma^2}}|x|} = 1 \pm (2u - 1), \quad 0 < u < 1;$$

$$\sqrt{2/\sigma^2}\,x = \ln[1 + (2u - 1)], \quad 0 < u < \frac{1}{2}, \quad x < 0;$$

$$\sqrt{2/\sigma^2}\,x = -\ln[1 - (2u - 1)], \quad \frac{1}{2} < u < 1, \quad x > 0,$$

d'où

$$\mathrm{sgn}\,x = \mathrm{sgn}\,(2u - 1). \qquad (3.161)$$

Ainsi, pour un bruit laplacien

$$f[F_1^{-1}(u)] = \sqrt{\frac{2}{\sigma^2}}\,\mathrm{sgn}\,(2u - 1). \qquad (3.162)$$

Considérons un bruit réparti suivant une loi logistique *) (de variance $\sigma^2$) de densité de probabilité

$$w_1(x;\,0) = \frac{\pi}{\sigma\sqrt{3}}\, e^{-\frac{\pi}{\sigma\sqrt{3}}x} / (1 + e^{-\frac{\pi}{\sigma\sqrt{3}}x})^2. \qquad (3.163)$$

Il est facile de voir que cette distribution est symétrique par rapport à l'origine. On a

$$\frac{-w_1'(x;\,0)}{w_1(x;\,0)} = \frac{\pi}{\sigma\sqrt{3}}\,\frac{e^{-\frac{\pi}{\sigma\sqrt{3}}x} - 1}{e^{-\frac{\pi}{\sigma\sqrt{3}}x} + 1}, \qquad (3.164)$$

$$F_1(x) = (1 + e^{-\frac{\pi x}{\sigma\sqrt{3}}})^{-1} = u, \qquad (3.164')$$

d'où

$$-\frac{\pi}{\sigma\sqrt{3}}\,x = \ln\frac{1-u}{u}, \quad 0 < u < 1. \qquad (3.164'')$$

---

*) La variance de la loi logistique $e^{-x}/(1 + e^{-x})^2$ est égale à

$$2\int_0^1 \ln^2 u\,\frac{du}{(1-u)^2} = \frac{2\pi^2}{6} = \frac{\pi^2}{3}.$$

En vertu de (3.158), (3.164) à (3.164″) pour un bruit logistique on a

$$f[F_1^{-1}(u)] = (2u-1)\frac{\pi}{\sigma\sqrt{3}}. \tag{3.165}$$

On peut maintenant, en utilisant (3.157), formuler les algorithmes de rang asymptotiquement optimaux de détection d'un signal déterministe noyé dans les bruits additifs répartis comme nous l'avons vu ci-dessus.

Dans le cas d'un bruit additif normal on prend la décision que le signal est présent si (voir (3.159))

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i F^{-1}\left(\frac{R_i}{n+1}\right) \geqslant C, \tag{3.166}$$

ce qui coïncide avec l'algorithme de rang de Van der Waerden (voir (2.134)).

Dans le cas d'un bruit additif laplacien on prend la décision que le signal est présent si (voir (3.162))

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i \operatorname{sgn}\left(R_i - \frac{n+1}{2}\right) \geqslant C, \tag{3.167}$$

ce qui coïncide avec l'algorithme des médianes (voir (2.132)).

Dans le cas d'un bruit additif logistique on prend la décision que le signal est présent si (voir (3.165))

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i R_i \geqslant C, \tag{3.168}$$

ce qui correspond à l'algorithme de rang de Wilcoxon (voir (2.133)).

Nous avons ainsi établi le fait que les algorithmes de rang envisagés dans le chapitre 2 de détection d'un signal déterministe sont asymptotiquement optimaux pour certains types de bruits additifs stationnaires indépendants.

**3.2.4. Algorithme asymptotiquement optimal signe-rang de détection d'un signal déterministe.** Lorsque l'on utilise les algorithmes signe-rang, on peut ne pas exiger que la composante constante soit absente dans le signal déterministe (voir 2.4.5).

Si la fonction de probabilité du bruit $w_1(x;\ 0)$ est symétrique par rapport à l'origine et satisfait à la condition (3.5), pour l'échantillon indépendant $x_1, \ldots, x_n$ appartenant à la distribution du bruit, on établit comme précédemment l'équivalence asymptotique

pour $n \to \infty$ de la statistique (3.154) et de la statistique signe-rang

$$z_n(R_1^+, \ldots, R_n^+) = \\ = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i u(x_i) f\left[F_1^{-1}\left(\frac{R_i^+}{2n+2} + \frac{1}{2}\right)\right], \qquad (3.169)$$

où $R^+ = (R_1^+, \ldots, R_n^+)$ est le vecteur des rangs positifs (voir 2.2.3). La faible restriction (3.145′) sur le signal reste, tandis que la restriction (3.145) sur l'absence de la composante constante est remplacée par la condition que les valeurs $|s_i|$ sont limitées pour $i$ quelconque. L'équivalence asymptotique a également lieu dans le cas où l'échantillon appartient à la distribution du mélange d'un signal $\lambda s(t)$ et d'un bruit pour $\lambda\sqrt{n} = \gamma$.

La règle asymptotiquement optimale de détection d'un signal déterministe, non paramétrique dans la classe des bruits à distribution symétrique, peut se formuler comme suit : on prend la décision que le signal est présent si

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i u(x_i) f\left[F_1^{-1}\left(\frac{R_i^+}{2n+2} + \frac{1}{2}\right)\right] \geqslant C, \qquad (3.170)$$

et la décision qu'il n'y a pas de signal si c'est l'inégalité inverse qui est vérifiée.

Considérons les algorithmes asymptotiquement optimaux de détection d'un signal constant $s(t) \equiv a > 0$ noyé dans un bruit additif de distribution symétrique. Dans ce cas on prend la décision que le signal est présent si

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} u(x_i) f\left[F_1^{-1}\left(\frac{R_i^+}{2n+2} + \frac{1}{2}\right)\right] \geqslant C \qquad (3.170')$$

En utilisant (3.162), (3.165) et (3.169) on peut maintenant proposer les algorithmes signe-rang asymptotiquement optimaux de détection d'un signal constant noyé dans des bruits additifs de certains types.

Pour un bruit additif normal on prend la décision que le signal est présent si (voir (3.159))

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} u(x_i) F^{-1}\left(\frac{R_i^+}{2n+2} + \frac{1}{2}\right) \geqslant C, \qquad (3.171)$$

ce qui coïncide avec l'algorithme signe-rang de Van der Waerden (2.103). Dans le cas d'un bruit additif laplacien, on décide que le

signal est présent si (voir (3.162))

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} u(x_i) \operatorname{sgn} \frac{R_i^+}{n+1} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u(x_i) \geqslant C, \qquad (3.171')$$

ce qui coïncide avec l'algorithme unilatéral de signe (2.47). Pour un bruit additif logistique, on prend la décision que le signal est présent si (voir (3.165))

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} u(x_i) R_i^+ \geqslant C, \qquad (3.171'')$$

ce qui coïncide avec l'algorithme signe-rang de Wilcoxon (2.91).

On a établi ainsi que les algorithmes signe-rang envisagés ci-dessus utilisés pour la détection d'un signal constant sont asymptotiquement optimaux pour certains types de bruits additifs: l'algorithme de Van der Waerden pour un bruit normal, l'algorithme de signe pour un bruit de Laplace, algorithme de Wilcoxon pour un bruit logistique.

**3.2.5. Stabilité de l'algorithme de rang asymptotiquement optimal de détection d'un signal déterministe.** Comme nous l'avons montré ci-dessus, lors de la détection d'un signal déterministe noyé dans un bruit à valeurs indépendantes on peut utiliser les algorithmes asymptotiquement optimaux de rang qui sont asymptotiquement équivalents aux algorithmes simples. On peut caractériser la stabilité de ces algorithmes par le coefficient d'efficacité asymptotique relative. Si pour la détection d'un signal déterministe noyé dans un bruit de distribution $u_1(x; 0)$ on utilise un algorithme asymptotiquement optimal de rang, « accordé » sur un bruit de distribution $w_1(x; 0)$, en reprenant les raisonnements de 3.1.5 on arrive à la conclusion que le coefficient d'efficacité relative de l'algorithme de rang, pour ce bruit « étranger », par rapport à l'algorithme asymptotiquement optimal est égal à (voir (3.42))

$$= \left( \int_0^1 f[F_1^{-1}(u)]\, g[G_1^{-1}(u)]\, du \right)^2 \times$$

$$\times \left[ \int_{-\infty}^{\infty} f^2(x)\, w_1(x; 0)\, dx \int_{-\infty}^{\infty} g^2(x)\, u_1(x; 0)\, dx \right]^{-1}, \qquad (3.172)$$

où $F_1^{-1}(x)$ et $G_1^{-1}(x)$ sont les fonctions inverses des fonctions de répartition $F_1(x)$ et $G_1(x)$ dont les dérivées sont $w_1(x; 0)$ et $u_1(x; 0)$ respectivement.

La formule (3.172) est symétrique par rapport aux distributions $w_1(x;0)$ et $u_1(x;0)$. Ceci signifie que l'efficacité relative d'un algorithme « accordé » sur le bruit $w_1(x;0)$ mais utilisé avec le bruit $u_1(x;0)$ est la même que celle d'un autre algorithme, « accordé » sur le bruit $u_1(x;0)$ mais utilisé avec le bruit $w_1(x;0)$. Ainsi, par exemple, l'efficacité relative d'un algorithme signe-rang « accordé » sur un bruit additif normal (voir (3.169)) mais utilisé avec un bruit laplacien est justement la même que celle d'un algorithme « accordé » sur un bruit additif de Laplace (voir (3.170)) lorsqu'on l'utilise dans un bruit normal. En vertu de (3.172), le coefficient d'efficacité asymptotique relative est dans ces cas

$$\rho = \frac{1}{\sigma^2}\left(\int_0^1 F^{-1}(u)\,\mathrm{sgn}\,(2u-1)\,du\right)^2 =$$

$$= \left(\frac{2}{\sigma^2\sqrt{2\pi}}\int_0^\infty z e^{-z^2/2\sigma^2}\,dz\right)^2 = \frac{2}{\pi}. \qquad (3.173)$$

**3.2.6. Comparaison des efficacités des algorithmes de rang et simples asymptotiquement optimaux de détection des signaux.** En comparant (3.173) avec (3.45) et (3.46) on arrive à des conclusions intéressantes. Les efficacités relatives des algorithmes asymptotiquement optimaux de rang et simple « accordés » sur un bruit laplacien mais utilisés avec un bruit normal sont les mêmes et égales à $2/\pi$. L'efficacité relative de l'algorithme asymptotiquement optimal de rang, « accordé » sur un bruit normal, lorsqu'on l'utilise dans un bruit laplacien est $8/\pi \approx 2{,}5$ fois plus grande que celle de l'algorithme simple « accordé » sur un bruit normal, mais employé pour la détection dans un bruit laplacien.

L'expression générale du coefficient d'efficacité relative d'un algorithme de rang asymptotiquement optimal par rapport à un algorithme simple, les deux étant « accordés » sur un bruit de distribution $w_1(x;0)$ et utilisés en présence d'un bruit « étranger » de distribution $u_1(x;0)$, se présente comme suit :

$$\rho = \frac{\left(\int_0^1 f[F_1^{-1}(z)]\,g[G_1^{-1}(z)]\,dz\right)^2 \int_{-\infty}^{\infty} f^2(x)\,u_1(x;0)\,dx}{\left(\int_{-\infty}^{\infty} f(x)\,g(x)\,u_1(x;0)\,dx\right)^2 \int_{-\infty}^{\infty} f^2(x)\,w_1(x;0)\,dx}. \qquad (3.174)$$

Ce coefficient est égal au rapport des coefficients calculés à l'aide des formules (3.43) et (3.172).

Il est montré dans [26] que le cas $\rho \geqslant 1$ a lieu lorsque la distribution $w_1(x;0)$ est celle d'un bruit additif normal. Autrement dit, la stabilité des algorithmes de rang « accordés » sur un bruit normal

est toujours supérieure à celle des algorithmes simples « accordés » également sur un bruit normal. Selon [26] cette propriété des algorithmes de rang apparaît pour une distribution quelconque du bruit additif.

## 3.3. ALGORITHMES ASYMPTOTIQUEMENT OPTIMAUX DE DÉTECTION DES SIGNAUX D'APRÈS DES OBSERVATIONS INDÉPENDANTES QUANTIFIÉES

**3.3.1. Quantification en amplitude.** Vu le large emploi des calculateurs numériques dans des systèmes informationnels modernes, il paraît indispensable d'étendre la théorie asymptotique exposée au § 3.2 des algorithmes asymptotiquement optimaux de détection des signaux au cas où les processus observés sont préalablement *quantifiés en amplitude.*

La quantification en amplitude consiste en ce qui suit. Le domaine des valeurs possibles de l'échantillon (sous-ensemble ou ensemble de toutes les valeurs réelles) est divisé en $m+1 \geqslant 2$ intervalles disjoints $E_k = (z_k, z_{k+1})$, $k = 0, \ldots, m$. Ainsi, la quantification est donnée par un vecteur à $m$ dimensions $\mathbf{z} = (z_1, \ldots, z_m)$, $z_0 = -\infty$, $z_{m+1} = \infty$.

A la sortie du dispositif de quantification, au lieu de l'échantillon observé $\mathbf{x} = (x_1, \ldots, x_m)$ on obtient un échantillon de même taille $n$ de grandeurs quantifiées $k^{(1)}, \ldots, k^{(n)}$, dont chaque élément ne peut prendre qu'une des valeurs de l'ensemble fini de $m+1$ valeurs, avec des probabilités

$$p_k = \int_{z_k}^{z_{k+1}} w_1(x;\ 0)\,dx; \quad k = 1, \ldots, m, \tag{3.175}$$

en l'absence de signal et

$$p_k(\lambda_n s_i) = \int_{z_k}^{z_{k+1}} w_1(x;\ \lambda_n s_i)\,dx; \quad k = 1, \ldots, m, \ i = 1, \ldots, n, \tag{3.176}$$

en présence du signal.

**3.3.2. Développement asymptotique du logarithme du rapport de vraisemblance pour des échantillons quantifiés.** Comme dans le cas des échantillons non quantifiés, la synthèse des algorithmes asymptotiquement optimaux pour les processus quantifiés se base sur le développement asymptotique du logarithme du rapport de vraisemblance des échantillons quantifiés.

Indiquons les conditions indispensables pour que le développement asymptotique mentionné existe. Soit $w_1(x;\ 0)$ la densité de

probabilité du bruit satisfaisant aux conditions (3.3) à (3.6), de plus la fonction $p_k(\vartheta)$ de (3.176) est dérivable par rapport à $\vartheta$.

La quantité d'information de Fisher pour les échantillons quantifiés est donnée par

$$I_z = \sum_{k=1}^{m} \left[ \frac{\partial}{\partial \vartheta} \ln p_k(\vartheta) |_{\vartheta=0} \right]^2 / p_k. \tag{3.177}$$

Supposons que $I_z < \infty$ et que la puissance du signal $s(t)$ soit limitée, c'est-à-dire que soit vérifiée la condition (3.9'). Alors pour $\lambda_n \sqrt{n} = \gamma$, où $0 < \gamma < \infty$ et $n \to \infty$, on a le développement asymptotique suivant du logarithme du rapport de vraisemblance $l_z$ pour des échantillons indépendants quantifiés:

$$\ln l_z(\mathrm{x}) = \sum_{i=1}^{n} \sum_{k=1}^{m} \chi_k(x_i) \ln \frac{p_k(\gamma s_i/\sqrt{n})}{p_k} =$$

$$= \frac{\gamma}{\sqrt{n}} \sum_{i=1}^{n} s_i \sum_{k=1}^{m} d_k \chi_k(x_i) - \frac{\gamma^2}{2} I_z W_s + \alpha_n(\mathrm{x}, \gamma s/\sqrt{n}), \tag{3.178}$$

où

$$d_k = \frac{\partial}{\partial \vartheta} \ln p_k(\vartheta) |_{\vartheta=0}, \tag{3.179}$$

$$\chi_k(x_i) = \begin{cases} 1, & x_i \in E_k, \\ 0, & x_i \bar{\in} E_k, \end{cases} \tag{3.180}$$

$$I_z = \sum_{k=1}^{m} p_k d_k^2, \quad W_s = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} s_i^2 \tag{3.180'}$$

(voir (3.9)). Dans le développement (3.178), le terme résiduel $\alpha_n$ tend en probabilité vers zéro pour $n \to \infty$, tant pour l'hypothèse $H$ que pour l'alternative $K$. Cette assertion se démontre à l'aide des résultats mentionnés dans 3.1.2 (voir [17]).

Remarquons que l'expression (3.178) peut être obtenue à partir de (3.8) en remplaçant $I_f$ par $I_z$ et la fonction $f(x_i)$ par la fonction

$$f_z(x_i) = \sum_{k=1}^{m} d_k \chi_k(x_i) = \sum_{k=1}^{m} \frac{\partial}{\partial \vartheta} \ln p_k(\vartheta) \Big|_{\vartheta=0} \chi_k(x_i). \tag{3.181}$$

Il est facile de voir que $m_1\{f_z(x) \mid H\} = 0$, $M_2\{f_z(x) \mid H\} = I_z$. En vertu du théorème de la limite centrale, le terme linéaire du développement (3.178)

$$y_{z_n}(\mathrm{x}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i f_z(x_i) \tag{3.182}$$

est réparti suivant une loi de probabilité *asymptotiquement normale* de paramètres $(0, I_z W_s)$ pour l'hypothèse $H$, et de paramètres $(\gamma I_z W_s, I_z W_s)$ pour l'alternative $K$. Lorsque les ensembles $E_k$ se rétrécissent en des points, à savoir $z_k \to z_{k+1}$, $\Delta z_k = z_{k+1} - z_k$, alors

$$p_k(\vartheta) \to w_1(z_k; \vartheta)\, \Delta z_k,\ f_z(x) \to f(x),\ I_z \to I_f \tag{3.183}$$

et la statistique (3.182) devient la statistique (3.9) pour un échantillon non quantifié.

Remarquons que pour $m$ fini quelconque, en vertu de (3.177) et de (3.5) on a

$$I_z \leqslant I_f, \tag{3.184}$$

c'est-à-dire que la quantité d'information de Fisher pour un échantillon quantifié n'est pas supérieure à cette quantité pour un échantillon non quantifié.

**3.3.3. Algorithme asymptotiquement optimal de détection d'un signal déterministe.** La règle asymptotiquement optimale de détection, d'après un échantillon quantifié d'observations, d'un signal déterministe noyé dans des bruits indépendants est

$$\begin{aligned} &\Phi(\mathbf{x}) = 1 \text{ (le signal est présent) si } y_{z_n}(\mathbf{x}) \geqslant C, \\ &\Phi(\mathbf{x}) = 0 \text{ (le signal est absent) si } y_{z_n}(\mathbf{x}) < C, \end{aligned} \tag{3.185}$$

où $y_{z_n}(\mathbf{x})$ est donné par (3.182).

Connaissant les valeurs des paramètres de la distribution limite pour l'hypothèse $H$ et vu que la statistique $y_{z_n}(\mathbf{x})$ est asymptotiquement normale, on trouve la valeur du seuil $C$ correspondant à une probabilité donnée $\alpha$ de fausse alarme

$$C = x_\alpha \sqrt{I_z W_s}, \tag{3.186}$$

où $x_\alpha$ est l'$\alpha$-quantile de la loi normale. En se servant des valeurs des paramètres de la distribution limite pour l'alternative, on établit une relation entre la constante $\gamma$ et les probabilités données de fausse alarme $\alpha$ et de perte du signal $\beta$ :

$$x_{1-\beta} = (C - \gamma I_z W_s)/\sqrt{I_z W_s}; \tag{3.187}$$

en réunissant (3.186) et (3.187) et en substituant la valeur de $C$ donnée par (3.186) on obtient

$$(x_\alpha - x_{1-\beta})^2 = \gamma^2 I_z W_s. \tag{3.188}$$

Lors de la quantification des observations, tout comme lors du rangement, une partie de l'information utile est en général perdue. Cependant, à la différence des algorithmes asymptotiquement optimaux de rang, qui sont aussi efficaces que les algorithmes simples,

les algorithmes asymptotiquement optimaux adaptés à des observations quantifiées accusent une réduction de l'efficacité. Il est donc intéressant de comparer l'efficacité de ces algorithmes selon qu'ils sont appelés à traiter des observations quantifiées ou non. Pour mesure de l'efficacité, on prend le coefficient d'efficacité asymptotique relative.

En utilisant (1.72) et les valeurs des paramètres des distributions limites des statistiques à comparer, on trouve que le coefficient d'efficacité asymptotique relative de l'algorithme asymptotiquement optimal (3.185) prévu pour les observations quantifiées, par rapport à l'algorithme (3.13′) utilisé pour les observations non quantifiées, soit

$$\rho = I_z/I_f, \tag{3.189}$$

est égal au rapport des quantités d'information de Fisher pour les échantillons quantifié et non quantifié (voir (3.5) et (3.177)). Compte tenu de (3.183), il est facile de voir que ce coefficient est non supérieur à l'unité. Il montre de combien de fois, s'il y a quantification des observations, il faut augmenter la taille de l'échantillon pour que l'algorithme de détection du signal d'après l'échantillon quantifié conserve les mêmes valeurs des probabilités d'erreurs $\alpha$ et $\beta$ que l'algorithme traitant des échantillons non quantifiés.

Si la quantification est telle que $I_z = I_f$, alors d'après (3.189) $\rho = 1$, c'est-à-dire que les algorithmes asymptotiquement optimaux de détection d'après des échantillons quantifiés et non quantifiés sont *équivalents*.

**3.3.4. Détection d'un signal constant.** En vertu de (3.182) et (3.184), l'algorithme asymptotiquement optimal de détection d'un signal déterministe d'après des échantillons quantifiés consiste à faire une somme des valeurs $s_i$ du signal, pondérées par les fonctions $f_z(x_i)$ (voir (3.181)). Pour un signal constant $s(t) \equiv a$, la statistique (3.182) s'écrit comme suit:

$$y_{z_n}(\mathbf{x}) = \frac{a}{\sqrt{n}} \sum_{i=1}^{n} f_z(x_i) = \frac{a}{\sqrt{n}} \sum_{k=1}^{m} d_k \sum_{i=1}^{n} \chi_k(x_i). \tag{3.190}$$

La somme $\sum_{i=1}^{n} \chi_k(x_i)$ est égale au nombre $n_k$ d'éléments dans l'échantillon $(x_1, \ldots, x_n)$ se trouvant sur l'intervalle $E_k$. Donc, omettant le facteur constant $a$, on peut écrire l'algorithme asymptotiquement optimal de détection d'un signal constant comme suit: le signal est présent si

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{m} d_k n_k \geqslant C, \quad \sum_{k=1}^{m} n_k = n, \tag{3.191}$$

et il n'y a pas de signal si c'est l'inégalité inverse qui est vérifiée. La valeur du seuil $C$ et la constante $\gamma$ sont données par (3.186) et (3.187) où il y a lieu de poser $W_s = 1$.

L'algorithme (3.191) est bien plus simple que l'algorithme général utilisant la statistique (3.182), car il est basé sur le vecteur $(n_1, \ldots, n_m)$ de dimension constante $m$, restant inchangée lorsqu'on augmente le nombre d'observations et ne dépendant que du nombre de niveaux de quantification.

Lorsqu'il s'agit de la détection d'un signal constant noyé dans un bruit de valeur moyenne nulle et dont la distribution est symétrique, le maximum de la quantité d'information $I_z$ d'après Fisher est atteint lors de la quantification en deux niveaux $E_1 = (-\infty, 0)$, $E_2 = (0, \infty)$. Dans ce cas en vertu de (3.175) et (3.176) on a

$$p_1 = p_2 = 1/2, \quad p_1(\vartheta) = \int_{-\infty}^{0} w_1(x-\vartheta)\,dx,$$

$$p_2(\vartheta) = \int_{0}^{\infty} w_1(x-\vartheta)\,dx$$

et en vertu de (3.179) on a

$$d_1 = -2w_1(0), \quad d_2 = 2w_1(0), \quad I_z = 4w_1^2(0). \tag{3.192}$$

En substituant (3.192) dans (3.191), on trouve l'algorithme asymptotiquement optimal de détection d'un signal noyé dans un bruit de distribution symétrique si l'on procède à une quantification en deux niveaux : le signal est présent si

$$\frac{1}{\sqrt{n}}(n_2 - n_1) = \frac{1}{\sqrt{n}}(n - 2n_1) \geqslant C, \tag{3.193}$$

où l'on reconnaît l'algorithme unilatéral de signe

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} u(x_i) \geqslant C. \tag{3.193'}$$

Remarquons que pour un bruit additif laplacien (voir (3.23)) $w_1^2(0) = \frac{1}{2\sigma^2}$ et par conséquent

$$I_z = 2/\sigma^2 = I_f, \tag{3.194}$$

i.e. on a la quantité d'information de Fisher pour des observations non quantifiées (voir (3.26)). Ceci signifie que dans ce cas les algorithmes asymptotiquement optimaux de détection d'après des échantillons quantifiés et non quantifiés sont équivalents. Le résultat obte-

nu n'est pas inattendu car l'algorithme asymptotiquement optimal de détection d'un signal constant noyé dans un bruit laplacien additif est justement un algorithme de signe (voir 3.1.3).

**3.3.5. Algorithme optimal simplifié de détection d'un signal déterministe d'après des observations quantifiées.** Comme nous l'avons montré dans le paragraphe précédent, l'algorithme asymptotiquement optimal de détection d'un signal d'après des observations quantifiées se simplifie beaucoup dans le cas où le signal est une grandeur constante sur l'intervalle d'observation. Il semble intéressant d'utiliser l'algorithme (3.191) pour un signal variable. Il est également intéressant d'estimer de combien baisse l'efficacité de cet algorithme simplifié par rapport à l'algorithme asymptotiquement optimal. La statistique (3.191) est asymptotiquement normale tant pour l'hypothèse $H$ que pour l'alternative $K$, les paramètres de la distribution limite pour l'hypothèse $H$ (absence de signal) étant $(0, \mathrm{I}_z)$. Le calcul des paramètres de la distribution limite pour l'alternative $K$ (présence d'un signal déterministe variable) est plus compliqué. On peut montrer (voir [2]) que la valeur moyenne est égale à

$$m_1 \left\{ \frac{1}{\sqrt{n}} \sum_{k=1}^{m} d_k n_k \mid K \right\} = \gamma \mathrm{I}_z a_s, \tag{3.195}$$

où

$$a_s = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} s_k, \tag{3.196}$$

et la variance est la même que pour l'hypothèse $H$.

Connaissant les valeurs des paramètres des distributions limites des statistiques, il est aisé de trouver, conformément à la formule générale (1.72), le coefficient d'efficacité asymptotique relative de l'algorithme optimal simplifié (3.191) de détection d'un signal déterministe par rapport à l'algorithme optimal (3.185):

$$\rho = a_s / \sqrt{\mathrm{W}_s}. \tag{3.197}$$

**3.3.6. Quantification optimale.** Lorsque l'on utilise des algorithmes de détection d'après des échantillons quantifiés, le problème est de trouver le quantificateur optimal, c'est-à-dire la partition $z_{\mathrm{opt}}$ telle que, pour un nombre donné $m$ de niveaux, on ait

$$\mathrm{I}_{z_{\mathrm{opt}}} = \sup_z \mathrm{I}_z. \tag{3.198}$$

Comme montré dans [2], la partition optimale $z_{\mathrm{opt}}$ satisfait au système d'équations

$$(d_k - d_{k-1})\,[d_k + d_{k-1} - 2f(z_k)] = 0; \quad k = 1, \ldots, m, \tag{3.199}$$

où $f(x)$ est donné par (3.4). Si $f(x)$ est une fonction monotone, le système (3.199) se simplifie et devient

$$d_k + d_{k-1} - 2f(z_k) = 0, \quad k = 1, \ldots, m. \tag{3.199'}$$

La résolution des systèmes (3.199), (3.199') n'est en général possible que par des méthodes numériques. On voit l'intérêt d'un développement, voisin de l'optimal, mais moins compliqué. Nous dirons que la suite des partitions $\mathbf{z}^*$ est asymptotiquement optimale, si

$$\lim_{m\to\infty} (I_f - I_{z^*})/(I_f - I_{z_{opt}}) = 1. \tag{3.200}$$

Pour certaines conditions (admissibles du point de vue pratique) $\mathbf{z}^*$ satisfait au système d'équations

$$\int_{-\infty}^{z_k^*} r(x)\,dx = \frac{k}{m+1} \int_{-\infty}^{\infty} r(x)\,dx; \quad k = 1, \ldots, m, \tag{3.201}$$

où

$$r(x) = \left[\left(\frac{\partial^2}{\partial x\,\partial\vartheta} \ln w_1(x;\ \vartheta)\,|_{\vartheta=0}\right)^2 \Big/ w_1(x;\ 0)\right]^{1/3}. \tag{3.201'}$$

Le coefficient d'efficacité asymptotique pour l'algorithme utilisant la partition $\mathbf{z}^*$ par rapport à l'algorithme asymptotiquement optimal est donné par l'expression suivante:

$$\rho = 1 - \frac{1}{12(m+1)^2 I_f}\left(\int_{-\infty}^{\infty} r(x)\,dx\right)^2 + o(1/m^2). \tag{3.202}$$

Dans le cas de la distribution de Rayleigh généralisée par exemple, dont la loi de probabilité est

$$w_1(x;\ \vartheta) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2-\vartheta}{2\sigma^2}\right) I_0\left(\frac{x\sqrt{\vartheta}}{\sigma}\right), \quad x > 0, \ \vartheta > 0,$$

la partition asymptotiquement optimale $\mathbf{z}^*$ est

$$z_k^* = \sigma\left(6 \ln \frac{m+1}{m+1-k}\right)^{\frac{1}{2}}; \quad k = 1, \ldots, m, \ z_0 = 0,$$

et

$$\rho = 1 - \frac{2{,}25}{(m+1)^2} + o(1/m^2). \tag{3.203}$$

## 3.4. STRUCTURE DES ALGORITHMES ASYMPTOTIQUEMENT OPTIMAUX DE DÉTECTION DES SIGNAUX D'APRÈS DES OBSERVATIONS CORRÉLÉES

**3.4.1. Modèle de bruits corrélés.** Tout comme dans le cas des observations indépendantes, la synthèse des algorithmes asymptotiquement optimaux de détection des signaux déterministes et quasi déterministes noyés dans des bruits corrélés est basée sur l'étude du développement asymptotique du logarithme du rapport de vraisemblance. Dans le cas des observations indépendantes, le logarithme du rapport de vraisemblance est égal à la somme des logarithmes du rapport de vraisemblance pour chaque observation. Pour certaines limitations sur la distribution unidimensionnelle du signal noyé dans un bruit, ceci a permis de trouver une règle assez simple de détection du signal. On peut également obtenir des résultats analogues dans le cas des observations corrélées à condition d'adopter un modèle de bruit pour lequel le rapport de vraisemblance se factorise, et donc le logarithme de vraisemblance est une somme de variables aléatoires. Tel est par exemple un processus markovien multiconnexe (voir [1.20]). De plus, un processus markovien multiconnexe est un modèle suffisamment général de bruits réels bien qu'il ne puisse prétendre à l'universalité.

**3.4.2. Développement asymptotique du logarithme du rapport de vraisemblance.** Soit $\mathbf{x}^n_{-k} = (\mathbf{x}_{-k}, \ldots, \mathbf{x}_n)$ un échantillon de la réalisation $x(t)$ du processus observé, $x_i = x(t_i)$, $t_i - t_{i-1} = \tau$, $i = -k, \ldots, n$. Le problème est de vérifier l'hypothèse $H$ selon laquelle l'échantillon appartient à la distribution $w(\mathbf{x}^n_{-k}; 0)$ contre l'alternative $K$ qu'il appartient à la distribution $w(\mathbf{x}^n_{-k}, \lambda_n \mathbf{s}^n_{-k})$, où $\mathbf{s}^n_{-k} = (s_{-k}, \ldots, s_n)$, $s_i = s(t_i)$ sont les valeurs de la fonction du signal à des instants donnés; $\lambda_n$ l'amplitude du signal. Soit $\mathbf{x}^n_{-k}$ une suite markovienne homogène $k$ fois connexe, stationnaire pour l'hypothèse $H$. Soit de plus

$$\lambda_n = \gamma/\sqrt{n}, \quad 0 < \gamma < \infty.$$

Compte tenu de la factorisation de la densité multidimensionnelle de la suite markovienne $k$ fois connexe (voir [1.20]) le logarithme du rapport de vraisemblance pour l'échantillon observé peut s'écrire comme suit:

$$\ln \frac{w(\mathbf{x}^n_{-k}; \gamma \mathbf{s}^n_{-k}/\sqrt{n})}{w(\mathbf{x}^n_{-k}; 0)} = \ln \frac{w(\mathbf{x}^{-1}_{-k}; \gamma \mathbf{s}^{-1}_{-k}/\sqrt{n})}{w(\mathbf{x}^{-1}_{-k}; 0)} + \\ + \sum_{i=0}^{n} \ln \frac{w(x_i \mid \mathbf{x}^{i-1}_{i-k}; \gamma \mathbf{s}^{i}_{i-k}/\sqrt{n})}{w(x_i \mid \mathbf{x}^{i-1}_{i-k}; 0)}, \qquad (3.204)$$

où $w(\mathbf{x}_{-k}^{-1}; \vartheta)$ est la densité de probabilité $k$-dimensionnelle de la suite markovienne $\mathbf{x}_{-k}^{n}$ et $w(x \mid \mathbf{y}; \vartheta)$ la densité de probabilité de la transition de $\mathbf{y}$ à $x$ ($\mathbf{y}$ étant un vecteur à $k$ dimensions). Ces densités dépendent du paramètre (du signal) à $k+1$ dimensions

$$\vartheta = \gamma s/\sqrt{n}.$$

Introduisons le vecteur $\mathbf{z} = (\mathbf{y}, x)$ à $k+1$ dimensions; imposons aux densités de probabilité mentionnées les conditions suivantes:

a) $w(\mathbf{z}; \vartheta)$ est continue en $\vartheta$ au point $\vartheta = 0$ uniformément sur $\forall \mathbf{z} \in Z$,

b) la densité de probabilité de transition admet le développement

$$w(x \mid \mathbf{y}; \vartheta) = w(x \mid \mathbf{y}; 0)\,[1 + \mathbf{f}'(\mathbf{z})\,\vartheta - |\vartheta|^2\,\delta(\mathbf{z}; \vartheta)], \quad (3.205)$$

$$m_1\{\mathbf{f}(\mathbf{z}) \mid \mathbf{y}, H\} = \int_{-\infty}^{\infty} \mathbf{f}(\mathbf{z})\, w(x \mid \mathbf{y}; 0)\, dx = 0, \quad (3.206)$$

$$m_1\{\mathbf{f}(\mathbf{z})\,\mathbf{f}'(\mathbf{z}) \mid H\} = \mathbf{K}, \ \det \mathbf{K} < \infty, \quad (3.207)$$

$$m_1\{\delta^2(\mathbf{z}; \vartheta) \mid H\} < D, \ |\vartheta| < R, \quad (3.208)$$

où $D$ et $R$ sont des constantes positives et $\mathbf{f}'$ est le vecteur transposé.

A partir de (3.206) on obtient également

$$m_1\{\mathbf{f}(\mathbf{z}) \mid H\} = m_1\{m_1\{\mathbf{f}(\mathbf{z}) \mid \mathbf{y}, H\}\} = 0. \quad (3.209)$$

Pour que le développement (3.205) existe, il suffit qu'au point $\vartheta = 0$ existent en moyenne quadratique les dérivées secondes mixtes de la densité de probabilité de transition, c'est-à-dire

$$m_2\left\{\frac{\partial^2}{\partial\vartheta_i\,\partial\vartheta_j}\, w(x \mid \mathbf{y}; \vartheta) \mid H\right\} < c, \quad |\vartheta| < R \quad (3.210)$$

(voir également [35]).

Pour les conditions mentionnées, le logarithme du rapport de vraisemblance admet le développement asymptotique suivant:

$$\ln l(\mathbf{x}_{-k}^{n}; \gamma \mathbf{s}_{-k}^{n}/\sqrt{n}) = \frac{\gamma}{\sqrt{n}} \sum_{i=0}^{n} \mathbf{f}'(\mathbf{x}_{i-k}^{i})\, \mathbf{s}_{i-k}^{i} - \frac{\gamma^2}{2}\,\mathrm{tr}\,[\mathbf{QK}] + \alpha_n(\mathbf{x}_{-k}^{n}; \gamma \mathbf{s}_{-k}^{n}/\sqrt{n}), \quad (3.211)$$

où Q et K sont les matrices définies positives égales à

$$\mathbf{Q}=(Q_{ij}),\ Q_{ij}=Q_{i-j},\qquad i,\ j=0,\ \ldots,k,$$

$$Q_l(\tau)=\lim_{n\to\infty}\frac{1}{n}\sum_{p=0}^{n}s_p s_{p+l}=\lim_{T\to\infty}\frac{1}{T}\int_0^T s(t)\,s(t+l\tau)\,dt, \quad (3.212)$$

$$\mathbf{K}=(K_{ij}),\ K_{ij}=m_1\{f_i f_j\,|\,H\},\quad i,\ j=0,\ \ldots,k, \quad (3.213)$$

et les composantes du vecteur $\mathbf{f}\,(\mathbf{x}^i_{i-k})$ sont *)

$$f_j(\mathbf{x}^i_{i-k})=\frac{\partial}{\partial s_{i-j}}\ln w\left(x_i\,|\,\mathbf{x}^{i-1}_{i-k};\ \gamma\mathbf{s}^i_{i-k}/\sqrt{n}\right)\Big|_{s^i_{i-k}=0},\qquad j=0,\ \ldots,k, \quad (3.214)$$

le terme résiduel $\alpha_n\to 0$ en probabilité lorsque $n\to\infty$ tant pour l'hypothèse $H$ que pour l'alternative $K$ si

a) des limites finies $Q_{ij}$ existent;

b) $m_4\{\mathbf{f}'(\mathbf{x})\,\mathbf{e}\mid H\}<\infty$, $\mathbf{e}$ étant le vecteur unité à $k+1$ dimensions;

c) $\operatorname{tr}[\mathbf{QK}]>0$.

Si le signal est périodique, les conditions a) et c) sont toujours vérifiées et la condition b) peut être affaiblie, à savoir, on peut exiger que

$$m_1\{[\mathbf{e}'\mathbf{f}(\mathbf{x})]^{2+\varepsilon}\,|\,H\}<\infty,\quad \varepsilon>0.$$

Pour prouver la légitimité du développement asymptotique (3.211) on procède comme suit: pour l'hypothèse $H$, on montre la convergence vers zéro du terme résiduel dans le développement du logarithme du rapport de vraisemblance en une série de Taylor, en gardant les termes linéaire et quadratique du développement. Puis on montre que la distribution du terme quadratique devient pour l'hypothèse $H$ une fonction delta au voisinage du point $(\gamma^2/2)\operatorname{tr}[\mathbf{QK}]$. On utilise ensuite la théorie de Le Cam (voir 3.1.2) pour démontrer l'équivalence asymptotique des suites de distributions lorsque les hypothèses $H$ et $K$ « se rapprochent ».

Le premier terme du développement asymptotique du logarithme du rapport de vraisemblance

$$Y_n(\mathbf{x}^n_{-k})=\frac{1}{\sqrt{n}}\sum_{i=0}^{n}\mathbf{f}'(\mathbf{x}^i_{i-k})\,\mathbf{s}^i_{i-k}=$$

$$=\frac{1}{\sqrt{n}}\sum_{i=0}^{n}\sum_{j=0}^{k}f_j(\mathbf{x}^i_{i-k})\,s(t_{i-k+j}) \quad (3.215)$$

*) Notons qu'en vertu de (3.214) on prend les dérivées par rapport à la valeur du signal à l'instant d'observation et par rapport aux valeurs du signal à $k$ instants précédents.

est une variable aléatoire asymptotiquement normale, de paramètres (0, tr [QK]) pour l'hypothèse $H$ et de paramètres ($\gamma$ tr [QK], tr [QK]) pour l'alternative $K$.

En toute rigueur il faudrait avoir dans (3.215) le facteur de normalisation $1/\sqrt{n+1}$. Il est évident cependant que dans un développement asymptotique on peut se satisfaire de $1/\sqrt{n}$.

Pour démontrer la normalisation de la somme de variables aléatoires *corrélées* dans le second membre de (3.215) pour l'hypothèse $H$ lorsque $n \to \infty$ on utilise le théorème de la limite centrale pour une somme de variables aléatoires « fortement mélangées », ayant en vue qu'une suite markovienne $k$ fois connexe est douée de cette propriété (voir [4]). La valeur moyenne de la statistique (3.215) est nulle pour l'hypothèse $H$ en vertu de (3.209). La variance de cette statistique pour l'hypothèse $H$ peut être facilement calculée si l'on met à profit le fait que les composantes des vecteurs aléatoires $f(\mathbf{x}^i_{i-k})$ ne sont pas corrélées deux à deux, à savoir

$$m_1\{f_l(\mathbf{x}^i_{i-k})\, f_g(\mathbf{x}^j_{j-k}) \mid H\} = K_{lg}\delta_{ij}, \tag{3.216}$$

où $\delta_{ij}$ est le symbole de Kronecker. En utilisant la théorie de Le Cam on peut montrer que la statistique (3.215) de paramètres ($\gamma$ tr [QK], tr [QK]) est asymptotiquement normale pour l'hypothèse $K$.

**3.4.3. Algorithmes asymptotiquement optimaux de détection d'un signal déterministe noyé dans un bruit markovien $k$ fois connexe.** En vertu de (3.211) et (3.215) l'algorithme asymptotiquement optimal de détection d'un signal déterministe noyé dans un bruit markovien $k$ fois connexe (pas obligatoirement additif) est

$$\begin{aligned} &\Phi(\mathbf{x}^n_{-k}) = 1 \text{ (le signal est présent) si } Y_n(\mathbf{x}^n_{-k}) \geqslant C, \\ &\Phi(\mathbf{x}^n_{-k}) = 0 \text{ (le signal est absent) si } Y_n(\mathbf{x}^n_{-k}) < C, \end{aligned} \tag{3.217}$$

$$m_1\{\Phi(\mathbf{x}^n_{-k}) \mid H\} = \alpha, \tag{3.218}$$

$$m_1\{\Phi(\mathbf{x}^n_{-k}) \mid K\} = 1 - \beta, \tag{3.219}$$

et la caractéristique de fonctionnement de l'algorithme asymptotiquement optimal (3.217) pour $n \to \infty$ tend vers la caractéristique de fonctionnement de l'algorithme optimal (suivant le critère de Neyman-Pearson).

La statistique $Y_n(\mathbf{x}^n_{-k})$ étant asymptotiquement normale, on obtient à partir de (3.218) une formule donnant la valeur du seuil d'après une probabilité donnée de fausse alarme $\alpha$, soit:

$$C = x_\alpha \sqrt{\operatorname{tr}[\mathbf{QK}]}, \tag{3.220}$$

où $x_\alpha$ est l' $\alpha$-quantile de la loi normale.

A partir de (3.219) on obtient également

$$x_{1-\beta} = x_\alpha - \gamma \sqrt{\operatorname{tr}[\mathbf{QK}]}, \tag{3.221}$$

d'où *)

$$\gamma = (x_\alpha - x_{1-\beta}) / \sqrt{\operatorname{tr}[\mathbf{QK}]}. \tag{3.222}$$

Remarquons que l'algorithme asymptotiquement optimal (3.13') de détection d'un signal déterministe noyé dans un bruit à valeurs lndépendantes s'obtient à partir de (3.217) comme un cas particuiier lorsque

$$w(x_i \mid \mathbf{x}_{i-k}^{i-1}, \mathbf{s}_{i-k}^{i}) \equiv w_1(x_i;\ s_i), \quad k = 0.$$

Dans ce cas le dispositif de détection n'a pas besoin d'élément inertiel (« mémoire ») tenant compte des observations et des valeurs antécédentes du signal. Il se compose d'un convertisseur non linéaire non inertiel des données d'entrée, de caractéristique

$$f(x) = \frac{d}{d\vartheta} \ln w_1(x;\ \vartheta)\,|_{\vartheta=0},$$

et d'un corrélateur donnant la corrélation mutuelle $\sum_{i=1}^{n} s_i f(x_i)$ du signal et des données observées transformées.

**3.4.4. Réalisation pratique de l'algorithme asymptotiquement optimal de détection d'un signal déterministe.** L'algorithme de détection (3.217) utilisant la statistique (3.215) préscrit l'exécution des opérations suivantes. Tout d'abord, on accumule dans la mémoire du calculateur $k$ valeurs échantillonnées du processus $\mathbf{x}_{-k}^{-1}$ observées aux instants $t_{-k}, \ldots, t_1$, et $k+1$ valeurs du signal déterministe $\mathbf{s}_{-k}^{0}$ aux mêmes instants et à l'instant $t_0$. Après une observation $x_0$ à l'instant $t_0$ on calcule la somme corrélationnelle de $k+1$ composantes des vecteurs $\mathbf{s}_{-k}^{0}$ et $\mathbf{f}(\mathbf{x}_{-k}^{0})$. Les composantes de ce dernier sont calculées d'après les densités de probabilité conditionnelles données des échantillons pour des valeurs $\mathbf{x}_{-k}^{-1}$, $\mathbf{s}_{-k}^{0}$ retenues dans la mémoire.

La somme corrélationnelle calculée est mémorisée de même que les vecteurs $\mathbf{x}_{-k+1}^{0}$, $\mathbf{s}_{-k+1}^{1}$ ; à l'instant $t_1$ on effectue l'observation suivante $x_1$. Puis on calcule une nouvelle somme corrélationnelle $\mathbf{f}'(\mathbf{x}_{-k+1}^{1})\mathbf{s}_{-k+1}^{1}$ et ainsi de suite $n+1$ fois pour l'échantillon observé $x_0, \ldots, x_n$. Enfin on totalise $n+1$ valeurs des sommes corrélationnelles et on compare le résultat avec un seuil. Le schéma fonctionnel du calculateur de la statistique (3.215) est donné sur la figure 3.5. A partir de la réalisation d'entrée $x(t)$, à l'aide d'une ligne à retard à $k+1$ positions, on forme successivement les vecteurs $\mathbf{x}_{i-k}^{i} = [x(t_{i-k}), \ldots, x(t_i)]$, $i = 0, \ldots, n$. Pour chaque $i$, la

*) D'une manière analogue à (3.18), l'expression (3.222) détermine pour des valeurs données $\alpha$, $\beta$ tr [QK] la valeur de seuil de l'amplitude du signal.

quantité gardée dans la position $j$ de la mémoire (ligne à retard) attaque le $j$-ième calculateur spécial de la composante $f_j(\mathbf{x}_{i-k}^i)$, $j = 0, \ldots, k$, conformément à la formule (3.214). A l'aide d'un oscillateur local du signal déterministe $s(t)$, pour chaque $i$ donné, il se forme dans la ligne à retard les vecteurs

$$\mathbf{s}_{i-k}^i = [s(t_{i-k}), \ldots, s(t_i)], \quad i = 0, \ldots, n.$$

Les données fournies par les calculateurs spéciaux et celles délivrées par le bloc des valeurs du signal attaquent le corrélateur dont

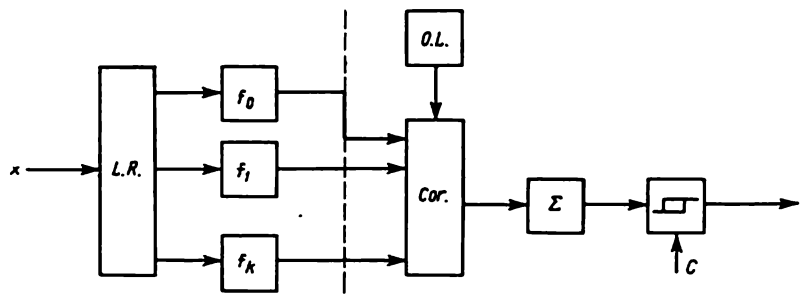



Fig. 3.5. Schéma fonctionnel du calculateur de la statistique $Y_n(\mathbf{x}_{-k}^n)$.

la sortie est reliée au sommateur-enregistreur des données. Le détecteur se simplifie dans le cas où l'on peut remplacer le traitement parallèle par un traitement séquentiel. Dans ce cas pour chaque $i$ donné le calculateur spécial $f_j$ effectue une dérivation, au total $k+1$ fois, en faisant avancer d'un pas le paramètre de dérivation. Le corrélateur sera dans ce cas également à un canal.

**3.4.5. Détection d'un signal déterministe noyé dans un bruit markovien.** Considérons le cas particulier de l'algorithme (3.217) lorsque le bruit est un processus markovien ($k = 1$) dont la densité de probabilité de transition est $w(x_i \mid x_{i-1}; 0)$. La statistique asymptotiquement normale (3.215) est dans ce cas

$$Y_n(\mathbf{x}_{-1}^n) = \frac{1}{\sqrt{n}} \sum_{i=0}^{n} \Big[ s_i \frac{\partial}{\partial s_{i-1}} \ln w(x_i \mid x_{i-1}; s_{i-1}, s_i)\big|_{s_{i-1}=s_i=0} +$$

$$+ s_{i-1} \frac{\partial}{\partial s_i} \ln w(x_i \mid x_{i-1}; s_{i-1}, s_i)\big|_{s_{i-1}=s_i=0} \Big]. \qquad (3.223)$$

La valeur moyenne et la variance de la statistique (3.223) pour l'hypothèse $H$ sont *)

$$m_1\{Y_n(\mathbf{x}_{-1}^n) \mid H\} = \frac{1}{\sqrt{n}} \sum_{i=0}^{n} \left[ s_i m_1 \left\{ \frac{\partial}{\partial s_{i-1}} \ln w(\cdot) \right\} + \right.$$

$$\left. + s_{i-1} m_1 \left\{ \frac{\partial}{\partial s_i} \ln w(\cdot) \right\} \right] = 0, \quad (3.224)$$

$$M_2\{Y_n(\mathbf{x}_{-1}^n) \mid H\} = \frac{1}{n} \sum_{i=0}^{n} \sum_{l=0}^{n} \left[ s_i s_l m_1 \left\{ \frac{\partial}{\partial s_{i-1}} \ln w(\cdot) \frac{\partial}{\partial s_{l-1}} \ln w(\cdot) \right\} + \right.$$

$$+ s_i s_{l-1} m_1 \left\{ \frac{\partial}{\partial s_{i-1}} \ln w(\cdot) \frac{\partial}{\partial s_l} \ln w(\cdot) \right\} +$$

$$+ s_{i-1} s_l m_1 \left\{ \frac{\partial}{\partial s_i} \ln w(\cdot) \frac{\partial}{\partial s_{l-1}} \ln w(\cdot) \right\} +$$

$$\left. + s_{i-1} s_{l-1} m_1 \left\{ \frac{\partial}{\partial s_i} \ln w(\cdot) \frac{\partial}{\partial s_l} \ln w(\cdot) \right\} \right]. \quad (3.225)$$

Mais compte tenu de (3.216), pour l'hypothèse $H$ on a

$$m_1 \left\{ \frac{\partial}{\partial s_i} \ln w(x_i \mid x_{i-1};\ s_{i-1},\ s_i) \frac{\partial}{\partial s_l} \ln w(x_l \mid x_{l-1};\ s_{l-1},\ s_l) \right\} = \delta_{il}. \quad (3.226)$$

A partir de (3.225) et (3.226) on trouve

$$M_2\{Y_n(\mathbf{x}_{-1}^n) \mid H\} = \frac{1}{n} \sum_{i=0}^{n} \left[ s_i^2 m_1 \left\{ \left[ \frac{\partial}{\partial s_{i-1}} \ln w(\cdot) \right]^2 \right\} + \right.$$

$$+ 2 s_i s_{i-1} m_1 \left\{ \frac{\partial}{\partial s_i} \ln w(\cdot) \frac{\partial}{\partial s_{i-1}} \ln w(\cdot) \right\} +$$

$$\left. + s_{i-1}^2 m_1 \left\{ \left[ \frac{\partial}{\partial s_i} \ln w(\cdot) \right]^2 \right\} \right] \quad (3.227)$$

et en utilisant (3.212), (3.213) on obtient **)

$$M_2\{Y_n(\mathbf{x}_{-1}^n) \mid H\} \sim Q_0 K_{00} + 2Q_1(\tau) K_{01} + Q_0 K_{11}, \quad (3.228)$$

---

*) Le point entre parenthèses signifie la répétition de l'argument de (3.223) et aussi que les dérivées sont prises pour les valeurs nulles de l'argument.

**) Il est évident que le second membre de (3.228) est la trace de la matrice QK (voir (3.211)).

où

$$Q_0 = \lim_{T\to\infty} \frac{1}{T} \int_0^T s^2(t)\,dt, \tag{3.229}$$

$$Q_1(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T s(t)\,s(t+\tau)\,dt, \tag{3.230}$$

$$K_{ij} = K_{ji} = m_1 \left\{ \frac{\partial}{\partial s_i} \ln w(x_i \mid x_{i-1};\ s_{i-1},\ s_i) \times \right.$$
$$\left. \times \frac{\partial}{\partial s_j} \ln w(x_j \mid x_{j-1};\ s_{j-1},\ s_j) \right\}_{s=0}; \quad i,\ j = 0;\ 1. \tag{3.231}$$

Lors de la détection d'un signal constant $s(t) \equiv a$ ($Q_0 = Q_1 = a^2$) noyé dans un bruit markovien normal additif pour lequel

$$\ln w(x_i \mid x_{i-1};\ a_{i-1},\ a_i) = -\frac{1}{2\sigma^2(1-r^2)} [(x_i - a_i)^2 - r(x_{i-1} - a_{i-1})^2] +$$
$$+ \ln [2\pi\sigma^2(1-r^2)]^{1/2}, \quad a_i = a_{i-1} = a, \tag{3.232}$$

on obtient :

$$f_0(\mathrm{x}_{i-1}^i) = x_i/[\sigma^2(1-r^2)], \quad f_1(\mathrm{x}_{i-1}^i) = -rx_{i-1}/[\sigma^2(1-r^2)], \tag{3.233}$$
$$K_{00} = [\sigma(1-r^2)]^{-2}, \quad K_{01} = r[\sigma(1-r^2)]^{-2}, \quad K_{11} = rK_{01} \tag{3.234}$$

et par conséquent

$$M_2\{Y_n(\mathrm{x}_{-1}^n) \mid H\} \sim \left(\frac{a}{\sigma}\right)^2 \Big/ (1-r)^2. \tag{3.235}$$

On prend la décision que le signal est présent, avec une probabilité donnée de fausse alarme $\alpha$, si (voir (3.220))

$$\frac{a}{\sigma^2(1-r^2)\sqrt{n}} \sum_{i=0}^{n} (x_i - rx_{i-1}) \geqslant x_\alpha \frac{a}{\sigma(1-r)} \tag{3.236}$$

ou

$$\frac{1}{n} \sum_{i=0}^{n} (x_i - rx_{i-1}) \geqslant x_\alpha \frac{\sigma(1+r)}{\sqrt{n}}. \tag{3.237}$$

L'inégalité (3.237) peut s'écrire comme suit :

$$\frac{1}{1-r} \frac{x_n - rx_{-1}}{n} + \frac{1}{n} \sum_{i=0}^{n-1} x_i \geqslant \frac{1+r}{1-r} \frac{x_\alpha \sigma}{\sqrt{n}}. \tag{3.238}$$

La règle de détection (3.238) coïncide avec la règle de détection d'un signal noyé dans un bruit normal markovien, optimale d'après le critère de Neyman-Pearson pour $n$ quelconque.

**3.4.6. Algorithme asymptotiquement optimal de détection d'un signal quasi déterministe noyé dans un bruit markovien multiplement connexe.** Considérons un signal quasi déterministe du type général (1.9). Le développement asymptotique du rapport de vraisemblance pour un vecteur $\vartheta$ donné est (voir (3.211))

$$l\left(\mathbf{x}_{-k}^{n}\left|\frac{\gamma}{\sqrt{n}}\vartheta'\varphi_{-k,\,n}\right.\right)\sim\exp\left\{\frac{\gamma}{\sqrt{n}}\vartheta'\sum_{i=0}^{n}\mathbf{f}'(\mathbf{x}_{i-k}^{i})\,\varphi_{i-k,\,i}-\frac{\gamma^2}{2}\vartheta'\mathbf{B}\vartheta\right\},\tag{3.239}$$

où $\varphi_{i-k,\,i}$ est une matrice de dimension $m\times(k+1)$, à éléments $\varphi_l(t_j)$, $l=1,\ldots,m$; $j\approx i-k,\ldots,i$; $\mathbf{B}$ est une matrice de dimension $m\times m$ à éléments $\operatorname{tr}[\mathbf{K}\mathbf{A}_{pq}]$, $p,\ q=1,\ldots,m$; $\mathbf{A}_{pq}$ est une matrice de dimension $(k+1)\times(k+1)$, dont les éléments $a_{pq}^{ij}$ sont

$$a_{pq}^{ij}=\lim_{T\to\infty}\frac{1}{T}\int_0^T\varphi_p(t+i\tau)\,\varphi_q(t+j\tau)\,dt,$$
$$i,\ j=0,\ldots,k;\ p,\ q=1,\ldots,m,\tag{3.240}$$

et $\mathbf{K}$ est une matrice de dimension $(k+1)\times(k+1)$ définie par (3.213). La statistique vectorielle (à $m$ dimensions)

$$\mathbf{Y}_n(\mathbf{x}_{-k}^{n})=\frac{1}{\sqrt{n}}\sum_{i=0}^{n}\mathbf{f}'(\mathbf{x}_{i-k}^{i})\,\varphi_{i-k,\,i}\tag{3.241}$$

de composantes

$$Y_{ln}(\mathbf{x}_{-k}^{n})=\frac{1}{\sqrt{n}}\sum_{i=0}^{n}\sum_{j=0}^{k}f_j(\mathbf{x}_{i-k}^{i})\,\varphi_l(t_{i-k+j})\tag{3.242}$$

est asymptotiquement normale tant pour l'hypothèse $H$ (il n'y a pas de signal) que pour l'alternative $K$ (le signal est présent), les matrices de corrélation des distributions normales limites coïncident pour l'hypothèse et pour l'alternative et sont égales à la matrice $\mathbf{B}$, le vecteur des moyennes est nul pour l'hypothèse $H$ et égal à $\gamma\,\mathbf{B}\vartheta$ pour l'alternative $K$.

Si l'on connaît la distribution mutuelle a priori $w_m(\vartheta)$ des paramètres $\vartheta$ du signal, le rapport de vraisemblance généralisé (moyenne sur $\vartheta$ du rapport de vraisemblance conditionnel) est égal à

$$\Lambda(\mathbf{Y}_n)=\int_{\Theta}w_m(\vartheta)\exp\left(\gamma\vartheta'\mathbf{Y}_n-\frac{\gamma^2}{2}\vartheta'\mathbf{B}\vartheta\right)d\vartheta.\tag{3.243}$$

En utilisant (3.243), on obtient l'algorithme asymptotiquement optimal de détection du signal quasi déterministe $s = \vartheta'\varphi$ suivant:

$$\begin{aligned} &\Phi[\mathbf{Y}_n(\mathbf{x}_{-k}^n)] = 1 \text{ (le signal est présent) si } \Lambda(\mathbf{Y}_n) \geqslant C, \\ &\Phi[\mathbf{Y}_n(\mathbf{x}_{-k}^n)] = 0 \text{ (il n'y a pas de signal) si } \Lambda(\mathbf{Y}_n) < C, \end{aligned} \tag{3.244}$$

avec

$$(2\pi)^{-m/2}\sqrt{\det \mathbf{B}^{-1}} \int_{\Theta} \int_{Y_m} \Phi(\mathbf{y}) \exp\left\{-\frac{1}{2}(\mathbf{y}-\gamma\mathbf{B}\vartheta)'\,\mathbf{B}^{-1}(\mathbf{y}-\gamma\mathbf{B}\vartheta)\right\} \times$$

$$\times\, w_m(\vartheta)\,d\mathbf{y}\,d\vartheta = \begin{cases} \alpha, & \gamma = 0, \\ 1-\beta, & \gamma > 0, \end{cases} \tag{3.244'}$$

où $\Theta$, $\mathbf{Y}_m$ sont des espaces euclidiens à $m$ dimensions.

Le système d'équations (3.244'), pour $\alpha$ et $\beta$ donnés, détermine le seuil $C$ et la constante $\gamma$.

Pour de nombreux signaux du type (1.9) qu'on rencontre dans la pratique, la règle (3.244) de choix d'une décision est uniformément asymptotiquement optimale, ce qui signifie que les caractéristiques de travail ne dépendent par de $\gamma$. Avec l'augmentation de la taille de l'échantillon ces caractéristiques se rapprochent des caractéristiques optimales (suivant le critère de Neyman-Pearson).

**3.4.7. Réalisation de l'algorithme asymptotiquement optimal de détection d'un signal quasi déterministe.** En vertu de (3.243) et (3.244), cet algorithme est doué de la propriété d'invariance partielle. Seuls le calculateur du vecteur de la statistique suffisante $\mathbf{Y}_n$ et la valeur du seuil dépendent du type de bruit et de son mode d'interaction avec le signal. L'autre bloc du détecteur, moyennant le rapport de vraisemblance, dépend seulement de la distribution des paramètres du signal (voir fig. 3.2).

Le schéma fonctionnel du calculateur des composantes du vecteur de la statistique suffisante est donné sur la figure 3.6. A partir de la réalisation d'entrée $x(t)$, la ligne de retard forme le vecteur $\mathbf{x}_{i-k} = [x(t_{i+k}), \ldots, x(t_i)]$. Les valeurs des composantes de ce vecteur sont utilisées dans $k+1$ calculateurs spéciaux donnant les quantités $f_j(\mathbf{x}_{i-k}^i)$, $j = 0, \ldots, k$, conformément à la formule (3.214). Cette partie générale du calculateur ne diffère pas de celle qui est donnée à gauche du pointillé sur la figure 3.5. Les $m$ oscillateurs locaux de fonctions de base $\varphi_l(t)$, $l = 1, \ldots, m$, servent à former, à l'aide de $m$ lignes de retard, les vecteurs $\boldsymbol{\varphi}_l$ de composantes $\varphi_l(t_{i-k}), \ldots, \varphi_l(t_i)$. Les données du $j$-ième calculateur spécial de $f_j$ attaquent $m$ corrélateurs identiques recevant également, en provenance des lignes de retard, les composantes des vecteurs de base. Les sorties de $k+1$ corrélateurs, correspondant chacun à une fonction de base donnée $\varphi_l$, sont sommées. Les calculs sont répétés pour

tous les $i = 0, \ldots, n$, puis les résultats sont totalisés dans l'enregistreur et sont mémorisés et, finalement, on obtient la composante $Y_{in}(\mathbf{x}^n_{-k})$.

Ainsi, le détecteur d'un signal quasi déterministe dépendant de $m$ paramètres aléatoires, noyé dans un bruit markovien $k$ fois connexe, se compose de $m$ calculateurs des composantes non linéaires du

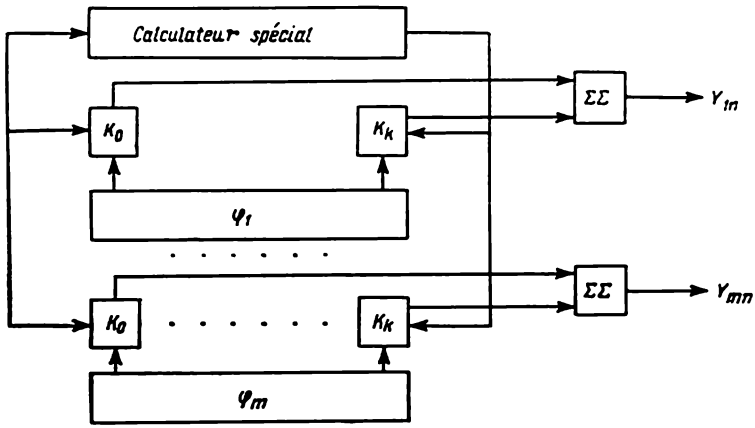



Fig. 3.6. Schéma fonctionnel du calculateur des composantes du vecteur de la statistique suffisante.

vecteur $\mathbf{Y}_n$, des corrélateurs du type discret, des sommateurs. La « mémoire » (lignes de retard) gardant $k+1$ valeurs des fonctions du signal et $k$ valeurs échantillonnées antérieures, est l'élément essentiel du système.

**3.4.8. Stabilité de l'algorithme asymptotiquement optimal de détection d'un signal quasi déterministe noyé dans un bruit markovien multiplement connexe.** Supposons que l'algorithme (3.244), asymptotiquement optimal pour un bruit distribué suivant $w(\mathbf{x}^n_{-k} \mid 0)$, soit utilisé dans des conditions où le bruit possède la distribution $u(\mathbf{x}^n_{-k} \mid 0)$. Pour caractériser la stabilité de l'algorithme (3.244) utilisé avec ce bruit « étranger », il y a lieu de trouver le coefficient d'efficacité asymptotique relative de cet algorithme par rapport à l'algorithme asymptotiquement optimal pour le bruit distribué suivant $u(\mathbf{x}^n_{-k} \mid 0)$. Ce problème est bien plus compliqué que le problème analogue étudié dans 3.1.5, où les bruits étaient à valeurs indépendantes. La difficulté réside en ce que maintenant il faut de nouveau étudier le comportement asymptotique de la statistique (3.241) pour un bruit « étranger », car à cause de la corrélation des composantes de (3.241), il n'est pas évident que la somme en est asymptoti-

quement normale. De plus, les paramètres de la distribution limite changent (vecteur des moyennes, matrice de corrélation), alors qu'on en a besoin pour résoudre le problème. Nous ne donnerons ici que la marche à suivre et les résultats finaux.

Tout d'abord il faut démontrer que la statistique (3.241) est asymptotiquement normale dans un bruit « étranger » pour l'hypothèse $H_u$, on trouve l'expression suivante pour les éléments de la matrice de corrélation $\mathbf{B}^{(u)}$ de la distribution limite:

$$B_{ij}^{(u)} = \lim_{n\to\infty} B_{ijn}^{(u)} = \lim_{n\to\infty} \left\{ \operatorname{tr}\left[ \mathbf{K}_0^{(u)} \frac{1}{n} \sum_{i=1}^{n} \mathbf{\Phi}_{lp}^{i,0} \right] + \right.$$

$$\left. + 2 \sum_{m=1}^{n} \operatorname{tr}\left[ \mathbf{K}_m^{(u)} \frac{1}{n} \sum_{i=m}^{n} \mathbf{\Phi}_{lp}^{im} \right] \right\}, \qquad (3.245)$$

où

$$\mathbf{K}_m^{(u)} = m_1 \{ \mathbf{f}'_w (\mathbf{x}_{i-k}^{i}) \mathbf{f}_w (\mathbf{x}_{i-k-m}^{i-m}) \mid H_u \}, \qquad (3.245')$$

$$\mathbf{\Phi}_{lp}^{im} = \varphi_l (\mathbf{t}_{i-k}^{i})' \varphi_p (\mathbf{t}_{i-k-m}^{i-m}), \qquad (3.245'')$$

$$\varphi_l (\mathbf{t}_{i-k}^{i}) = [\varphi_l (t_{i-k}), \ldots, \varphi_l (t_i)]. \qquad (3.245''')$$

Le vecteur des moyennes est nul si la condition

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \varphi_l (t_i) = 0, \quad l = 1, \ldots, m,$$

est vérifiée.

Puis on montre que la statistique (3.241) est asymptotiquement normale dans un bruit « étranger » pour l'alternative $K_u$. La matrice de corrélation de la distribution limite est la même que pour l'hypothèse $H_u$, tandis que le vecteur des moyennes est égal à $\gamma \mathbf{B}^{(u,w)} \mathbf{\Phi}$, où $\mathbf{B}^{(u,w)}$ est une matrice dont les éléments sont

$$B_{qr}^{(u,w)} = \lim_{n\to\infty} \left\{ \operatorname{tr}\left[ \mathbf{K}_0^{(u,w)} \frac{1}{n} \sum_{i=1}^{n} \mathbf{\Phi}_{qr}^{i,0} \right] + \right.$$

$$\left. + 2 \sum_{m=1}^{n} \operatorname{tr}\left[ \mathbf{K}_m^{(u,w)} \frac{1}{n} \sum_{i=m}^{n} \mathbf{\Phi}_{qr}^{i,m} \right] \right\}, \qquad (3.246)$$

où

$$(\mathbf{K}_m^{(u,w)})' = m_1 \{ \mathbf{f}'_u (\mathbf{x}_{i-k}^{i}) \mathbf{f}_w (\mathbf{x}_{i-k-m}^{i-m}) \mid H_u \}. \qquad (3.246')$$

Le coefficient cherché d'efficacité asymptotique relative est $\rho = (\gamma_0/\gamma_0^*)^2$ (voir (3.244')), où $\gamma_0$ et $\gamma_0^*$ sont les racines minimales

des équations

$$(2\pi)^{-m/2}\sqrt{\det \mathbf{B}^{-1}}\int\limits_{\Theta}\int\limits_{Y_m}\Phi(\mathbf{y})\exp\left\{-\frac{1}{2}(\mathbf{y}-\gamma\mathbf{B}\vartheta)'\,\mathbf{B}^{-1}(\mathbf{y}-\gamma\mathbf{B}\vartheta)\right\}\times$$
$$\times\, w_m(\vartheta)\,d\vartheta\,d\mathbf{y}=1-\beta_0,$$

$$(2\pi)^{m/2}\sqrt{\det \mathbf{B}^{(u,w)}}\int\limits_{\Theta}\int\limits_{Y_m}\Phi(\mathbf{y})\exp\left\{-\frac{1}{2}(\mathbf{y}-\gamma\mathbf{B}^{(u,w)}\vartheta)'\times\right.$$
$$\left.\times\,[\mathbf{B}^{(u,w)}]^{-1}(\mathbf{y}-\gamma\mathbf{B}^{(u,w)}\vartheta)\right\}w_m(\vartheta)\,d\vartheta\,d\mathbf{y}=1-\beta_0.$$

**3.4.9. Algorithme asympotiquement optimal de détection d'un signal modulé de phase aléatoire.** Considérons le signal à bande étroite (3.58). En utilisant (3.60), (3.60′) dans le cas où l'enveloppe $a(t)$ et la phase $\psi(t)$ du signal varient lentement par rapport à $\cos\omega_0 t$, on trouve

$$\lim_{T\to\infty}\frac{1}{T}\int\limits_0^T\varphi_1(t)\,\varphi_1(t+\tau)\,dt=\lim_{T\to\infty}\frac{1}{T}\int\limits_0^T\varphi_2(t)\,\varphi_2(t+\tau)\,dt=$$
$$=\frac{1}{2}\lim_{T\to\infty}\frac{1}{T}\int\limits_0^T a(t)\,a(t+\tau)=\frac{1}{2}B_a(\tau),\qquad(3.247)$$

$$\lim_{T\to\infty}\frac{1}{T}\int\limits_0^T\varphi_1(t)\,\varphi_2(t+\tau)\,dt=\lim_{T\to\infty}\frac{1}{T}\int\limits_0^T\varphi_2(t)\,\varphi_1(t+\tau)\,dt=0.\qquad(3.247')$$

En vertu de (3.247) et (3.247′), dans le cas envisagé la matrice $\mathbf{B}$ de dimension $2\times 2$ est diagonale, les éléments diagonaux étant

$$b=\frac{1}{2}\sum_{l=0}^{k}\sum_{j=0}^{k}K_{lj}B_a[(l-j)\tau].\qquad(3.248)$$

Ecrivons maintenant le rapport de vraisemblance généralisé sous la condition que la phase aléatoire $\varphi_0$ est répartie uniformément sur l'intervalle $(0,\ 2\pi)$:

$$\Lambda(\mathbf{Y}_n)=\int\limits_{-1}^{1}\int\limits_{-1}^{1}\frac{1}{\pi\sqrt{1-\vartheta_1^2}}\,\delta\left(\vartheta_2-\sqrt{1-\vartheta_1^2}\right)\times$$
$$\times\exp\left(\gamma\vartheta_1 Y_{1n}+\gamma\vartheta_2 Y_{2n}-\frac{\gamma^2}{2}b\right)d\vartheta_1\,d\vartheta_2=$$
$$=e^{-\gamma^2 b/2}\frac{1}{2\pi}\int\limits_0^{2\pi}\exp(\gamma Y_{1n}\cos\varphi+\gamma Y_{2n}\sin\varphi)\,d\varphi=$$
$$=\mathrm{I}_0(\gamma\,|\,Y_n|)e^{-\gamma^2 b/2}.\qquad(3.249)$$

En vertu de (3.249), le rapport de vraisemblance généralisé est une fonction monotone de la statistique

$$|\mathbf{Y}_n(\mathbf{x}_{-k}^n)|^2 = \frac{1}{n}\Big[\sum_{i=0}^{n} \mathbf{f}'(\mathbf{x}_{i-k}^i)\,\varphi_1(\mathbf{t}_{i-k}^i)\Big]^2 + \\ + \frac{1}{n}\Big[\sum_{i=0}^{n} \mathbf{f}'(\mathbf{x}_{i-k}^i)\,\varphi_2(\mathbf{t}_{i-k}^i)\Big]^2, \qquad (3.250)$$

où

$$\varphi_1(\mathbf{t}_{i-k}^i) = [a(t_{i-k})\cos[\omega_0 t_{i-k} + \psi(t_{i-k})], \ldots \\ \ldots, a(t_i)\cos[\omega_0 t_i + \psi(t_i)]], \qquad (3.251)$$

$$\varphi_2(\mathbf{t}_{i-k}^i) = [a(t_{i-k})\sin[\omega_0 t_{i-k} + \psi(t_{i-k})], \ldots \\ \ldots, a(t_i)\sin[\omega_0 t_i + \psi(t_i)]]. \qquad (3.252)$$

La statistique (3.250) peut s'écrire comme suit :

$$|\mathbf{Y}_n(\mathbf{x}_{-k}^n)|^2 = \frac{1}{n}\Big\{\sum_{i=0}^{n}\sum_{j=0}^{k} f_j(\mathbf{x}_{i-k}^i)\,a(t_{i-j})\cos[\omega_0 t_{i-j} + \psi(t_{i-j})]\Big\}^2 + \\ + \frac{1}{n}\Big\{\sum_{i=0}^{n}\sum_{j=0}^{k} f_j(\mathbf{x}_{i-k}^i)\,a(t_{i-j})\sin[\omega_0 t_{i-j} + \psi(t_{i-j})]\Big\}^2. \qquad (3.253)$$

L'algorithme asymptotiquement optimal de détection d'un signal quasi déterministe du type (3.58) de phase aléatoire uniformément répartie consiste à comparer la statistique (3.253) à un seuil déterminé par une probabilité donnée de fausse alarme $\alpha$ pour $\gamma$ quelconque (c'est-à-dire que cet algorithme est uniformément asymptotiquement optimal). Comme pour l'hypothèse $H$, les quantités $Y_{1n}$ et $Y_{2n}$ sont des variables aléatoires normales indépendantes de valeurs moyennes nulles et de variance $b$, le seuil mentionné est donné par

$$C_0 = b\chi_\alpha^2(2), \qquad (3.254)$$

où $\chi_\alpha^2(2)$ est l'$\alpha$-quantile de la loi du $\chi^2$ à deux degrés de liberté, c'est-à-dire de la loi exponentielle.

Pour l'hypothèse $K$, $Y_{1n}$ et $Y_{2n}$ sont des variables aléatoires normales indépendantes de valeurs moyennes $\gamma b$ et de variances $b$. La probabilité de perte du signal $\beta$ est alors donnée par l'égalité

$$b\chi_{1-\beta}^2(2, \delta) = C_0, \qquad (3.255)$$

où $\chi_{1-\beta}^2(2, \delta)$ est l'1-$\beta$-quantile de la loi du $\chi^2$ à deux degrés de liberté de paramètre d'écart $\delta = \gamma^2 b$.

Le dispositif de détection non cohérente d'un signal modulé à bande étroite de phase aléatoire noyé dans un bruit markovien multiplement connexe, contient deux canaux en quadrature où sont calculées les statistiques asymptotiquement suffisantes, correspondant l'une à la composante en phase du signal et l'autre à celle en quadrature. Puis comme toujours on a des détecteurs quadratiques, des sommateurs et un dispositif à seuil.

**3.4.10. Distribution normale des paramètres d'un signal.** Supposons que la distribution a priori $w_m(\vartheta)$ des paramètres d'un signal quasi déterministe du type (1.9) soit normale de vecteur des moyennes $\vartheta_0$ et de matrice de corrélation $\mathbf{D}$. A partir de (1.33) on voit que dans ce cas le rapport de vraisemblance généralisé (3.243) dépend d'une manière monotone de la statistique

$$\Psi(\mathbf{Y}_n) = \mathbf{Y}_n'\mathbf{B}^{-1}\mathbf{Y}_n - (\mathbf{Y}_n - \gamma\mathbf{B}\vartheta_0)'(\mathbf{B} + \gamma^2\mathbf{B}'\mathbf{D}^{-1}\mathbf{B})^{-1}(\mathbf{Y}_n - \gamma\mathbf{B}\vartheta_0), \quad (3.256)$$

et par conséquent la règle de choix d'une décision peut se formuler comme suit : le signal est présent si $\Psi(\mathbf{Y}_n) \geqslant C'$, le seuil $C'$ et la constante $\gamma$ sont déterminés pour $\alpha$ et $\beta$ donnés, à partir du système d'équations

$$\alpha = P\{\Psi(\mathbf{Y}_n) \geqslant C' \mid H\}, \quad (3.257)$$

$$1 - \beta = P\{\Psi(\mathbf{Y}_n) \geqslant C' \mid K\}. \quad (3.258)$$

Supposons que les fonctions d'autocorrélation et de corrélation mutuelle temporelles de la base $\{\varphi(t)\}$ sont telles que

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \varphi_p(t)\,\varphi_p(t+\tau)\,dt = B_\varphi(\tau), \quad (3.259)$$

i.e. $B_\varphi(\tau)$ ne dépendent pas de $p$, et pour $p \neq q$

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \varphi_p(t)\,\varphi_q(t+\tau)\,dt = B_{pq}(\tau) = B_{qp}(-\tau) = -B_{qp}(\tau), \quad (3.260)$$

i.e. $B_{qp}(\tau)$ est une fonction impaire de $\tau$. La matrice $\mathbf{B}$ est alors diagonale

$$\mathbf{B} = b\mathbf{I},\ b > 0, \quad (3.261)$$

$$b = \sum_{l=0}^{k} \sum_{j=0}^{k} K_{lj} B_\varphi[(l-j)\tau]. \quad (3.262)$$

Comme on sait, on peut toujours décorréler un ensemble de variables aléatoires $\vartheta$ de vecteur des moyennes $\vartheta_0$ et de matrice de corrélation $\mathbf{D}$, en effectuant sur $\vartheta$ une transformation linéaire

$$\eta = \vartheta'\mathbf{F}, \quad (3.263)$$

où F est une matrice orthogonale de vecteurs lignes $F_i$, $i = 1, \ldots, m$, satisfaisant à l'équation

$$\mathbf{D}\varphi = \mathbf{d}\varphi, \quad (\varphi_i, \varphi_j) = \delta_{ij}. \tag{3.264}$$

L'ensemble des variables aléatoires normales $\eta$ est indépendant, de vecteur des moyennes $\vartheta_0^* = \mathbf{F}\vartheta_0$ et de matrice de corrélation diagonale à éléments $d_i > 0$, $i = 1, \ldots, m$. La transformation orthogonale (3.263) des paramètres du signal signifie que le signal (1.9) est rapporté à une nouvelle base, soit

$$s(t) = \vartheta'\varphi(t) = \eta\mathbf{F}^{-1}\varphi(t) = \eta\varphi^*(t), \tag{3.265}$$

avec

$$\varphi^*(t) = \mathbf{F}^{-1}\varphi(t). \tag{3.266}$$

Il est facile de montrer que la transformation orthogonale de la base ne touche pas les relations (3.259), (3.260) si elle est vraie pour la base $\varphi(t)$. Ainsi, après transformation orthogonale de la base la statistique (3.256) devient

$$\Psi(\mathbf{Y}_n) = \mathbf{Y}_n'\mathbf{Y}_n/b - (\mathbf{Y}_n - \gamma b\vartheta_0^*)'(b\mathbf{I} + \gamma^2 b^2 \mathbf{d}^{-1})^{-1} \times$$

$$\times (\mathbf{Y}_n - \gamma b\vartheta_0^*) = \gamma^2 \sum_{j=1}^{m} (Y_{jn} + d_j\vartheta_{0j}^*/\gamma)^2 \frac{1}{d_j + \gamma^2 b} + \text{const}, \tag{3.267}$$

$$\vartheta_{0j}^* = \sum_{i=1}^{m} F_{ij}\vartheta_{0i}, \tag{3.268}$$

En vertu de (3.267), on a la règle asymptotiquement optimale suivante : le signal est présent si

$$\gamma^2 \sum_{j=1}^{m} (Y_{jn} + d_j\vartheta_{0j}^*/\gamma)^2/(d_j + \gamma^2 b) \geqslant C. \tag{3.269}$$

Le seuil $C$ et la constante $\gamma$ se trouvent à partir du système d'équations

$$P\left\{\gamma^2 \sum_{j=1}^{m} (Y_{jn} + d_j\vartheta_{0j}^*/\gamma)^2/(d_j + \gamma^2 b) \geqslant C \mid H\right\} = \alpha, \tag{3.270}$$

$$P\left\{\gamma^2 \sum_{j=1}^{m} (Y_{jn} + d_j\vartheta_{0j}^*/\gamma)^2/(d_j + \gamma^2 b) \geqslant C \mid K\right\} = 1 - \beta, \tag{3.271}$$

où

$$Y_{jn} = \frac{1}{\sqrt{n}} \sum_{i=0}^{n} \sum_{l=0}^{k} f_l(\mathbf{x}_{i-k}^i)\, \varphi_j(t_{i-k+l})$$

sont des variables aléatoires normales indépendantes de variances $b$ et dont les moyennes sont nulles pour l'hypothèse $H$ et égales à $\gamma b\vartheta^*_{0j}$ pour l'alternative $K$.

Le calcul des constantes $C$ et $\gamma$ revient à trouver la distribution de la somme finie pondérée des carrés de variables aléatoires normales indépendantes de moyennes et de variances différentes. Ces calculs aboutissent à une loi du $\chi^2$ non centrée tabulée à $m$ degrés de liberté

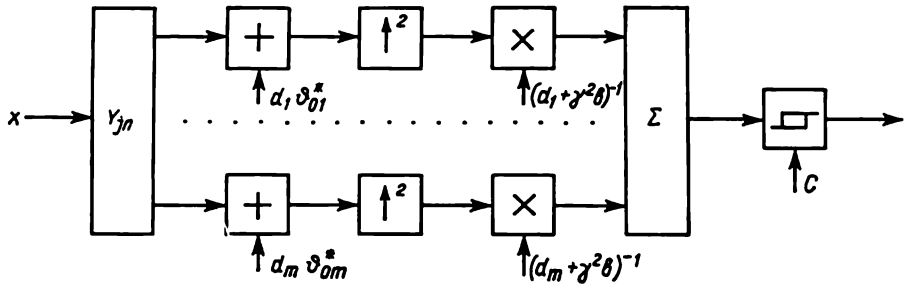



Fig. 3.7. Schéma fonctionnel du dispositif réalisant l'algorithme (3.269).

si $d_i = d$, c'est-à-dire seulement dans le cas d'une distribution des paramètres à symétrie sphérique. Dans ce cas, l'inégalité (3.269) peut s'écrire comme suit :

$$\frac{1}{b}\sum_{j=1}^{m}(Y_{jn}+d\vartheta_{0j}^*/\gamma)^2 \geqslant C. \tag{3.272}$$

Le paramètre d'écart de la statistique (3.272) est

$$\delta_0(\gamma)=\frac{d^2}{\gamma^2 b}|\vartheta_0^*|^2 \tag{3.273}$$

pour l'hypothèse $H$, et

$$\delta_1(\gamma)=\frac{(d+\gamma^2 b)^2}{\gamma^2 b}|\vartheta_0^*|^2 \tag{3.274}$$

pour l'alternative $K$, où

$$\vartheta_0^*|^2=\sum_{j=0}^{m}\vartheta_{0j}^{*2}. \tag{3.275}$$

Par conséquent, en vertu de (3.270) à (3.272), la constante $\gamma$ et le seuil $C$ pour $\alpha$ et $\beta$ donnés se trouvent à partir du système d'équations

$$\chi_\alpha^2[m,\ \delta_0(\gamma)]=\chi_{1-\beta}^2[m,\ \delta_1(\gamma)], \tag{3.276}$$

$$\chi_\alpha^2[m,\ \delta_0(\gamma)]=C. \tag{3.277}$$

Si $|\vartheta_0^*|=0$ la statistique (3.272) est répartie suivant une loi du $\chi^2$ centrée à $m$ degrés de liberté. Dans ce cas l'algorithme de dé-

tection est uniformément asymptotiquement optimal. Le seuil $C$ ne dépend pas de $\gamma$ et est égal à $\chi^2_\alpha(m)$. La réalisation pratique de l'algorithme (2.369) est assez simple (fig. 3.7). Les sorties des calculateurs spéciaux des composantes du vecteur $\mathbf{Y}_n$ sont sommées avec des constantes données, les sommes attaquent les dispositifs d'élévation au carré et les atténuateurs, les résultats étant aiguillés vers le sommateur général. La valeur de la somme est comparée au seuil.

**3.4.11. Détection d'un signal stochastique.** Les résultats du paragraphe précédent peuvent être utilisés pour établir l'algorithme asymptotiquement optimal de détection d'un signal stochastique, qui représente un processus aléatoire normal de moyenne nulle et de fonction de corrélation $B(t, y)$, noyé dans un bruit markovien multiplement connexe. Dans ce cas il y a lieu d'utiliser le développement orthogonal (1.12), la base $\{\varphi(t)\}$ étant déterminée à partir d'une équation intégrale homogène linéaire de noyau $B(t, y)$. Le signal stochastique a la même structure que le signal quasi déterministe que nous avons vu plus haut, à condition de poser $m = \infty$. L'ensemble des paramètres $\vartheta_k/\sqrt{\lambda_k}$, $k=1, 2, \ldots$, est un ensemble de variables aléatoires normales indépendantes de moyennes nulles et de variances $1/\lambda_k$. Dans ce cas on tire de (3.269) la règle asymptotiquement optimale suivante : le signal est présent si

$$\gamma^2 b \sum_{j=1}^{\infty} Y_{jn}^2/(\lambda_j+\gamma^2 b) \geqslant C \tag{3.278}$$

ou encore

$$\frac{\gamma^2 b}{n} \sum_{j=1}^{\infty} \Big[ \sum_{i=0}^{n} \sum_{l=0}^{k} f_l(\mathbf{x}_{i-k}^i)\, \varphi_l(t_{i-k+l}) \Big]^2 \Big/ (\lambda_j+\gamma^2 b) \geqslant C. \tag{3.279}$$

Introduisons la désignation

$$h(t, y) = \gamma^2 b \sum_{j=1}^{\infty} \varphi_j(t)\, \varphi_j(y)/(\lambda_j+\gamma^2 b). \tag{3.280}$$

Il est facile de voir que la fonction $h(t, y)$ satisfait à l'équation intégrale

$$\int_{-T}^{T} h(y, v)\, B(y, t)\, dy + h(t, v)/(\gamma^2 b) = B(t, v), \quad |t| \leqslant T, \quad |v| \leqslant T. \tag{3.281}$$

En utilisant (3.280) on peut écrire comme suit l'algorithme (3.278) :

$$\frac{1}{n} \sum_{i_1=0}^{n} \sum_{i_2=0}^{n} \sum_{p=0}^{k} \sum_{q=0}^{k} h(t_{i_1-p}, t_{i_2-q})\, f_p(\mathbf{x}_{i_1-k}^{i_1})\, f_q(\mathbf{x}_{i_2-k}^{i_2}) \geqslant C. \tag{3.282}$$

**3.4.12. Algorithmes de rang asymptotiquement optimaux de détection des signaux noyés dans des bruits corrélés.** Les algorithmes de rang non paramétriques de détection des signaux noyés dans des bruits indépendants ont été étudiés en détail au chapitre 2. Dans 2.7 nous avons noté que les propriétés non paramétriques de ces algorithmes cessent d'être non paramétriques si l'on les utilise avec des observations corrélées. Cependant les algorithmes de rang ont la propriété de maintenir constante la probabilité de fausse alarme pour une classe de processus obtenue à partir d'un processus donné, à l'aide d'une transformation non linéaire non inertielle monotone, même lorsque l'on utilise des observations corrélées.

L'étude de la distribution des statistiques de rang pour des échantillons corrélés de taille finie est, dans le cas général, un problème assez ardu (voir (2.43')). Nous allons utiliser l'équivalence asymptotique etudiée dans 3.2.1 des statistiques de rang et non ordonnées pour obtenir, à partir de (3.217), un algorithme de rang asymptotiquement optimal de détection d'un signal déterministe noyé dans un bruit markovien $k$ fois connexe, comme nous avons déjà déduit, pour un échantillon indépendant, l'algorithme (3.157) à partir de l'algorithme (3.13). A cet effet il y a lieu de vérifier que les relations (3.141) et (3.148) sont vérifiées pour une suite markovienne $k$ fois connexe. Ceci est vrai en vertu des résultats suivants obtenus dans [36].

Considérons un processus aléatoire stationnaire $x(t)$ satisfaisant à la condition de mélange intense, de coefficient de mélange $\alpha(k)$, limité par la condition (voir [4], page 388)

$$\sum_{k=1}^{\infty} k\sqrt{\alpha(k)} < \infty. \tag{3.283}$$

Servons-nous de l'échantillon $\mathbf{x}_1^n = (x_1, \ldots, x_n)$, $x_i = x(t_i)$ du processus $x(t)$ pour former la suite discrète

$$z_n\left(\frac{k}{n+1}\right) = \sqrt{n}\left[x^{(k)} - F_1^{-1}\left(\frac{k}{n+1}\right)\right], \quad k = 1, \ldots, n, \tag{3.284}$$

où $F_1(x)$ est la fonction de répartition du processus; $x^{(k)}$ la statistique d'ordre $k$ dans l'échantillon $\mathbf{x}_1^n$. Par interpolation linéaire, on obtient à partir de (3.284) le processus $z_n(t)$, continu sur l'intervalle $(0, 1)$. Supposons que pour un $\varepsilon > 0$ petit quelconque on ait

$$z_n^{\varepsilon}(t) = z_n(t), \quad t \in (\varepsilon, 1-\varepsilon),$$
$$z^{\varepsilon}(t) = z(t), \quad t \in (\varepsilon, 1-\varepsilon).$$

Dans [36] on montre qu'il existe un processus aléatoire normal $z(t)$ de moyenne nulle et de fonction de corrélation

$$B_z^{\varepsilon}(t, y) = m_1\{z(t), z(y)\},$$

qui entraîne pour $n \to \infty$ la convergence (faible) de $z_n^\varepsilon(t)$ vers $z^\varepsilon(t)$, la condition (3.283) étant suffisante pour que $|B_z(t, y)| < \infty$. Il en découle, en vertu de (3.284), la possibilité d'approximation de la valeur échantillonnée $x_k$ par $F_1^{-1}\left(\frac{R_k}{n+1}\right)$, ceci d'une manière d'autant plus précise que la taille $n$ de l'échantillon est plus grande. Une suite markovienne $k$ fois connexe étant un processus à mélange intense et donc satisfaisant à la condition (3.283), la dernière affirmation est vraie pour cette suite.

L'équivalence de la statistique simple $Y_n(\mathbf{x}_{-k}^n)$ (3.215) et de la statistique de rang

$$z_n(\mathbf{R}_{-k}^n) = \frac{1}{\sqrt{n}} \sum_{i=0}^{n} \mathbf{f}'\left[F_1^{-1}\left(\frac{R_{i-k}}{n+k+2}\right), \ldots, F_1^{-1}\left(\frac{R_i}{n+k+2}\right)\right] \mathbf{s}_{i-k}^i \tag{3.285}$$

s'écrit comme suit :

$$Y_n(\mathbf{x}_{-k}^n) = z_n(\mathbf{R}_{-k}^n) + \chi_n(\mathbf{x}_{-k}^n), \tag{3.286}$$

où $\chi_n(\mathbf{x}_{-k}^n) \to 0$ en probabilité pour $n \to \infty$.
En utilisant l'approximation mentionnée ci-dessus, on peut montrer que pour une suite markovienne $k$ fois connexe, la fonction $f$ étant suffisamment lisse et à condition que

$$\sup_n \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i \right| < \infty, \tag{3.287}$$

la suite des variables aléatoires $\chi_n$ dans (3.286) converge en probabilité vers zéro lorsque la taille de l'échantillon augmente indéfiniment.

Ainsi, l'égalité (3.286), généralisant (3.148) au cas d'un bruit corrélé, permet d'utiliser les résultats de 3.4.3 pour la synthèse d'un algorithme de rang asymptotiquement optimal de détection d'un signal noyé dans un bruit markovien $k$ fois connexe [1]. Une généralisation au cas des signaux quasi déterministes découle directement des résultats de 3.4.6.

**3.4.13. Algorithme analogique asymptotiquement optimal de discrimination des signaux.** Dans ce chapitre nous avons exposé d'une manière assez détaillée la théorie des algorithmes asymptotiquement optimaux discrets de détection des signaux d'après des observations indépendantes ou des observations markoviennes multiplement connexes. Pour conclure, indiquons une des applications possibles du principe d'optimum asymptotique dans les problèmes de discrimination des signaux « voisins » d'après des observations des réalisations continues [7].

Considérons un processus de diffusion $x_t = x(t)$, décrit par l'équation différentielle stochastique [19]

$$dx_t = s(\vartheta, t, x_t)\,dt + dW_t, \quad 0 \leqslant t \leqslant T, \tag{3.288}$$

où $W_t$ est un processus de Wiener; $s(\cdot)$ une fonction donnée; $\vartheta$ un paramètre du signal. On vérifie l'hypothèse $H$ que $\vartheta = \vartheta_0$ contre l'alternative $K$ que

$$\vartheta = \vartheta_0 + \varphi_T u, \quad u > 0, \tag{3.289}$$

avec $\varphi_T \to 0$ pour $T \to \infty$. Il est évident que $\varphi_T$ est un homologue de la suite $1/\sqrt{n}$ pour des échantillons discrets. Pour éviter les singularités, tout comme dans le cas discret, il faut imposer une condition sur la vitesse de convergence vers zéro de la grandeur $\varphi_T$ lorsque la durée de l'observation augmente indéfiniment. On arrive ainsi à la condition suivante: pour l'hypothèse $H$, la limite en probabilité est

$$\lim_{T\to\infty} \varphi_T^2 \int_0^T \dot{s}^2(\vartheta_0, t, x_t)\,dt = 1, \tag{3.290}$$

où

$$\dot{s}^2(\vartheta_0, t, x_t) = \left[\frac{\partial}{\partial\vartheta} s(\vartheta, t, x_t)\big|_{\vartheta=\vartheta_0}\right]^2. \tag{3.290'}$$

Désignons la réalisation $x_t$ sur l'intervalle $0 \leqslant t \leqslant T$ par $x^T$ et introduisons la statistique

$$y(x^T) = \varphi_T\left[\int_0^T \dot{s}(\vartheta_0, t, x_t)\,dx - \int_0^T \dot{s}(\vartheta_0, t, x_t)\,s(\vartheta_0, t, x_t)\,dt\right], \tag{3.291}$$

où l'intégrale stochastique dans (3.291) est entendue dans le sens que lui attribue Ito.

On peut montrer que le logarithme du rapport de vraisemblance pour le problème de la discrimination des signaux (vérification de l'hypothèse $H$ contre l'alternative $K$) admet le développement suivant (comparer avec (3.211)):

$$\ln l(x^T) = u y(x^T) - \frac{1}{2} u^2 + \psi_T(u, x^T), \tag{3.292}$$

$\psi_T$ converge en probabilité vers zéro pour $T \to \infty$ tant pour l'hypothèse que pour l'alternative. La statistique $y(x^T)$ est asymptotiquement suffisante et possède une distribution asymptotiquement normale de paramètres (0, 1) pour l'hypothèse $H$ et de paramètres $(u, 1)$ pour l'alternative $K$.

Pour que ce résultat fondamental ait lieu, il suffit que soit vérifiée la condition (3.290) et que $\ddot{s}(\vartheta, t, x_t)$ soit bornée.

En vertu de (3.292), l'algorithme asymptotiquement normal de discrimination des signaux est le suivant: pour une probabilité donnée $\alpha$ d'erreur de première espèce, on adopte l'alternative $K$ si

$$y(x^T) \geqslant x_\alpha, \tag{3.293}$$

et l'hypothèse $H$ si l'on a l'inégalité inverse de (3.293). $x_\alpha$ est ici l'$\alpha$-quantile de la loi normale.

A titre d'exemple simple, nous allons envisager le problème de la détection d'un signal stochastique ergodique, noyé dans un bruit blanc normal additif. Les observations sont décrites par l'équation différentielle stochastique suivante:

$$dx_t = \lambda s(x_t)\,dt + dW_t, \quad 0 \leqslant t \leqslant T. \tag{3.294}$$

On vérifie l'hypothèse $H$: $\lambda = 0$ (il n'y a pas de signal) contre l'alternative $K$: $\lambda > 0$ (le signal est présent). Pour $T \to \infty$, l'amplitude $\lambda$ du signal tend vers zéro, en remarquant que *)

$$\lambda = u\varphi_T. \tag{3.295}$$

Si le processus $x_t$ est ergodique, on a avec une probabilité égale à l'unité

$$\frac{1}{T}\int_0^T s^2(x_t)\,dt \to W_s < \infty, \tag{3.296}$$

par conséquent, en vertu de (3.290) on a

$$\varphi_T = 1/\sqrt{W_s T}. \tag{3.297}$$

On déduit de (3.291) et (3.293) la règle asymptotiquement optimale suivante de détection d'un signal: pour une probabilité donnée de fausse alarme $\alpha$ on adopte la décision que le signal est présent si

$$\frac{1}{\sqrt{W_s T}}\int_0^T s(x_t)\,dx_t \geqslant x_\alpha, \tag{3.298}$$

et la décision qu'il n'y a pas de signal si c'est l'inégalité inverse de (3.298) qui est vérifiée.

## 3.5. PROBLÈMES

3.1. Montrer que la caractéristique d'un convertisseur non linéaire non inertiel placé à l'entrée d'un détecteur asymptotiquement optimal d'un signal déterministe noyé dans un bruit additif est

$$f(x) = \operatorname{sgn} x/(1 - |x|) \tag{1}$$

*) Avec les désignations de (3.289) on a $\vartheta_0 = 0$, $\vartheta = \lambda$.

pour un bruit de densité de probabilité $w_1(x) = 1 - |x|$, $|x| \leqslant 1$;

$$f(x) = e^x - 1 \tag{2}$$

pour un bruit de densité de probabilité $w_1(x) = \exp(x - e^x)$;

$$f(x) = 2 \operatorname{sgn} x / (1 + |x|) \tag{3}$$

pour un bruit de densité de probabilité $w_1(x) = 0{,}5\,(1 + |x|)^{-2}$;

$$f(x) = (1 - e^{-x})/(1 + e^{-x}) \tag{4}$$

pour un bruit de distribution logistique (3.163).

3.2. Montrer que la caractéristique d'un convertisseur non linéaire non inertiel des rangs, à prévoir dans un détecteur asymptotiquement optimal de rang d'un signal déterministe noyé dans un bruit additif est

$$\Delta(u) = \operatorname{sgn}(2u - 1)/(1 - |2u - 1|)^{1/2} \tag{5}$$

pour un bruit de densité de probabilité $w_1(x) = 1 - |x|$, $|x| \leqslant 1$;

$$\Delta(u) = \ln \frac{1}{1-u} - 1 \tag{6}$$

pour un bruit de densité de probabilité $w_1(x) = \exp(x - e^x)$;

$$\Delta(u) = 2\,(1 - |2u - 1|) \operatorname{sgn}(2u - 1) \tag{7}$$

pour un bruit de densité de probabilité $w_1(x) = 0{,}5\,(1 + |x|)^{-2}$.

3.3. Montrer que l'algorithme, *optimal d'après le critère de Neyman-Pearson*, de détection d'un signal constant $s(t) \equiv a$ noyé dans un bruit additif de moyenne nulle, de variance $\sigma^2$ et distribué suivant la loi

$$w_1(x) = b_1 \exp[-(b_2 |x|)^k], \; k > 0, \tag{8}$$

où

$$b_1 = \frac{k\,[\Gamma(3/k)\,\Gamma(2/k)]^{1/2}}{2\sigma\Gamma(1/k)}, \quad b_2 = \frac{1}{\sigma}[\Gamma(3/k)\Gamma(1/k)]^{1/2} \tag{9}$$

d'après un échantillon indépendant $x_1, \ldots, x_n$ se formule comme suit : on adopte la décision que le signal est présent si

$$\sum_{i=1}^{n} (|x_i|^k - |x_i - a|^k) \geqslant C, \tag{10}$$

où le seuil $C$ est déterminé par une probabilité donnée de fausse alarme.

3.4. Pour les conditions du problème 3.3 montrer que l'algorithme *asymptotiquement optimal* de détection du signal est : on adopte la

décision que le signal est présent si

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} |x_i|^{k-1} \operatorname{sgn} x_i \geqslant C. \tag{11}$$

Examiner les cas particuliers de la distribution de Laplace ($k = 1$) et de la distribution normale ($k = 2$).

**3.5.** En utilisant (3.42) montrer que le coefficient d'efficacité asymptotique relative de l'algorithme (3.13) par rapport à l'algorithme linéaire, en présence d'un bruit additif (« étranger ») de densité de probabilité $u_1(x;\ 0)$ est

$$\rho = \sigma^2 \left[ \int_{-\infty}^{\infty} f'(x)\, u_1(x;\ 0)\, dx \right]^2 \left[ \int_{-\infty}^{\infty} f^2(x)\, u_1(x;\ 0)\, dx - \right.$$

$$\left. - \left( \int_{-\infty}^{\infty} f(x)\, u_1(x;\ 0)\, dx \right)^2 \right]^{-1}, \tag{12}$$

où $\sigma^2$ est la variance du bruit. Vérifier que dans le cas où en qualité de convertisseur non linéaire on utilise un limiteur parfait et si la distribution du bruit est symétrique, la formule (12) coïncide avec (2.70).

**3.6.** Pour les conditions du problème 3.5 montrer que le maximum de la fonctionnelle $\rho\ \{f(x)\}$ est

$$\max_f \rho = \sigma^2 \int_{-\infty}^{\infty} \frac{[u_1'(x;\ 0)]^2}{u_1(x;\ 0)}\, dx. \tag{13}$$

Vérifier que pour un bruit laplacien on a $\max\limits_f \rho = 2$.

**3.7.** Montrer que le coefficient d'efficacité asymptotique relative de l'algorithme de signe par rapport à l'algorithme linéaire, en présence d'un bruit additif de distribution (8), est

$$\rho = k^2 \Gamma(3/k)\, \Gamma^3(1/k). \tag{14}$$

**3.8.** En utilisant la formule (14), trouver l'expression suivante de $\max\limits_f \rho$ pour un bruit distribué suivant (8)

$$\max_f \rho = k^2 \Gamma(3/k)\, \Gamma(2 - 1/k)\, \Gamma^2(1/k). \tag{15}$$

Analyser pour différentes valeurs de $k$ l'expression

$$\max_f \rho/\rho = \Gamma\left(2 - \frac{1}{k}\right) \Gamma\left(\frac{1}{k}\right) = \left(1 - \frac{1}{k}\right)\left(2 - \frac{1}{k}\right) \pi/\sin\frac{\pi}{k}. \tag{16}$$

**3.9.** [17]. Soient $x_{c1}, \ldots, x_{cn}$ et $x_{s1}, \ldots, x_{sn}$ des observations indépendantes à $n$ instants des composantes indépendantes en qua-

drature du processus aléatoire à bande étroite

$$x(t) = x_c(t)\cos\omega_0 t + x_s(t)\sin\omega_0 t, \tag{17}$$

qui est soit un bruit à bande étroite, soit un mélange avec le signal $\lambda s(t)$, où

$$s(t) = a(t)\cos[\omega_0 t + \psi(t) - \varphi_0]. \tag{18}$$

Montrer que les algorithmes asymptotiquement optimaux analogo-discrets de détection des signaux envisagés sont basés sur les statistiques suivantes:

pour $\varphi_0 = 0$ (détection cohérente)

$$y_n(\mathbf{x}_c, \mathbf{x}_s) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} a_i[\cos\psi_i f(x_{ci}) + \sin\psi_i f(x_{si})], \tag{19}$$

où $a_i = a(t_i)$, $\psi_i = \psi(t_i)$; $f(x)$ est la caractéristique de l'élément non linéaire déterminée par la distribution du bruit (3.4);

pour une phase aléatoire uniformément répartie (détection non cohérente), on a

$$y_n(\mathbf{x}_c, \mathbf{x}_s) = \frac{1}{n}\left\{\left(\sum_{i=1}^{n} a_i[f(x_{ci})\cos\psi_i + f(x_{si})\sin\psi_i]\right)^2 + \right.$$
$$\left. + \left(\sum_{i=1}^{n} a_i[f(x_{si})\cos\psi_i - f(x_{ci})\cos\psi_c]\right)^2\right\}. \tag{20}$$

**3.10.** Montrer que dans les conditions du problème 3.9 les algorithmes numériques (non paramétriques) asymptotiquement optimaux de détection du signal (18), avec une quantification en deux niveaux, sont basés sur les statistiques suivantes:

pour $\varphi_0 = 0$

$$y_n(\mathbf{x}_c, \mathbf{x}_s) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} a_i(\cos\psi_i \operatorname{sgn} x_{ci} + \sin\psi_i \operatorname{sgn} x_{si}); \tag{21}$$

pour une phase aléatoire uniformément répartie

$$y_n(\mathbf{x}_c, \mathbf{x}_s) = \frac{1}{n}\left\{\left[\sum_{i=1}^{n} a_i(\cos\psi_i \operatorname{sgn} x_{ci} + \sin\psi_c \operatorname{sgn} x_{si})\right]^2 + \right.$$
$$\left. + \left[\sum_{i=1}^{n} a_i(\cos\psi_i \operatorname{sgn} x_{si} - \sin\psi_i \operatorname{sgn} x_{ci})\right]^2\right\}. \tag{22}$$

**3.11.** [17]. Montrer que l'algorithme numérique de phase asymptotiquement optimal de détection du signal à bande étroite (18)

noyé dans un bruit à bande étroite lorsque la phase est uniformément quantifiée en $m$ niveaux, est basé sur la statistique suivante (voir 3.3.2):

pour $\varphi_0 = 0$

$$y_n(\vartheta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} a_i \Big[\cos\psi_i \sum_{k=1}^{m} \chi_k(\vartheta_i) \cos\frac{\pi}{m} \times$$

$$\times (2k-1) + \sin\psi_i \sum_{k=1}^{m} \chi_k(\vartheta_i) \sin\frac{\pi}{m}(2k-1)\Big]; \qquad (23)$$

pour une phase aléatoire uniformément répartie

$$y_n(\vartheta) = \frac{1}{n}\Big\{\Big(\sum_{i=1}^{n} a_i \Big[\cos\psi_i \sum_{k=1}^{m} \chi_k(\vartheta_i) \cos\frac{\pi}{m}(2k-1) +$$

$$+ \sin\psi_i \sum_{k=1}^{m} \chi_k(\vartheta_i) \sin\frac{\pi}{m}(2k-1)\Big]\Big)^2 +$$

$$+ \Big(\sum_{i=1}^{n} a_i \Big[\cos\psi_i \sum_{k=1}^{m} \chi_k(\vartheta_i) \sin\frac{\pi}{m}(2k-1) +$$

$$+ \sin\psi_i \sum_{k=1}^{m} \chi_k(\vartheta_i) \cos\frac{\pi}{m}(2k-1)\Big]\Big)^2\Big\}. \qquad (24)$$

3.12. Montrer que pour un signal déterministe à bande étroite manipulé en phase tel que $a_i = 1$ si $\cos\psi_i = 0$, et $a_i = -1$ si $\cos\psi_i = -1$, l'algorithme asymptotiquement optimal de détection d'un signal noyé dans un bruit à bande étroite consiste à comparer avec un seuil les statistiques suivantes:

algorithme de phase (analogo-discret)

$$y_n(\vartheta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_i \cos\vartheta_i, \qquad (25)$$

où

$$h_i = \begin{cases} +1 & \text{pour } \cos\psi_i = 0, \\ -1 & \text{pour } \cos\psi_i = -1, \end{cases}$$

algorithme de phase avec quantification uniforme de la phase

$$y_n(\vartheta) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_i \sum_{k=1}^{m} \chi_k(\vartheta_i) \frac{\sin\varphi_{k+1} - \sin\varphi_k}{\varphi_{k+1} - \varphi_k}, \qquad (26)$$

où $\varphi_k = (2\pi/m)(k-1)$.

Montrer que le coefficient d'efficacité asymptotique relative de l'algorithme (26) par rapport à l'algorithme (25) est égal à

$$\rho = \frac{1}{\pi} \sum_{k=1}^{m} (\sin \varphi_{k+1} - \sin \varphi_k)^2 / (\varphi_{k+1} - \varphi_k). \tag{27}$$

Vérifier que pour $m = 4$

$$\rho = 8/\pi^2.$$

**3.13.** Montrer que l'algorithme asymptotiquement optimal de détection d'un signal quasi déterministe, dont les paramètres sont un ensemble de $m$ variables aléatoires normales indépendantes de moyennes nulles, dans un mélange avec un bruit à valeurs indépendantes préscrit la comparaison au seuil $C$ de la statistique

$$\chi^2(m) = \frac{1}{I_f} |y_n|^2, \tag{28}$$

où $y_n$ est donnée par (3.47). On suppose que les fonctions de base $\varphi_k(t)$ du signal, figurant dans l'expression (3.47), sont asymptotiquement orthogonales et normées. Obtenir l'expression du seuil

$$C = \chi^2_\alpha(m), \tag{29}$$

où $\chi^2_\alpha(m)$ est l'$\alpha$-quantile de la loi du $\chi^2$ centrée. (A remarquer que la formule (29) ne contient pas la constante $\gamma$.)

**3.14.** Montrer que pour la détection dans un bruit d'un signal quasi déterministe $\lambda s(t)$, où

$$s(t) = \sum_{i=1}^{m} \vartheta_i \varphi_i(t) \tag{30}$$

et $\vartheta_1, \ldots, \vartheta_m$ sont les paramètres aléatoires de densité de probabilité mutuelle $w_m(\vartheta)$, l'algorithme asymptotiquement optimal, utilisant des échantillons indépendants quantifiés, préscrit la comparaison à un seuil de la statistique

$$\Phi(Y_n) = \int_{\Theta} \exp\left[\gamma \vartheta' Y_n - \frac{\gamma^2}{2} I_z \vartheta' A \vartheta\right] w_m(\vartheta)\, d\vartheta, \tag{31}$$

où

$$Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi(t_i) \sum_{k=0}^{m} a_k \chi_k(x_i), \tag{32}$$

$$A = (a_{pq}),\ a_{pq} = \lim_{T \to \infty} \frac{1}{T} \int_0^T \varphi_p(t)\, \varphi_q(t)\, dt, \tag{33}$$

les autres désignations étant données dans 3.3.

CHAPITRE 4

# ALGORITHMES ADAPTATIFS

## 4.1. CLASSIFICATION DES OBSERVATIONS DANS LE CAS DES DISTRIBUTIONS NORMALES

**4.1.1. Algorithme de classification avec apprentissage.** De nombreux problèmes posés par la radiodétection, les télécommunications et la gestion avec indétermination a priori se formulent en termes de la théorie de la classification des observations (reconnaissance des images). Le problème consiste alors à rapporter l'objet observé à une des classes, dont la description probabiliste complète est inconnue. Dans ce cas on ne peut pas utiliser les résultats classiques de la théorie des décisions statistiques. Pour trouver la solution on peut faire usage des observations étalons (échantillons d'apprentissage). A partir des échantillons d'apprentissage on formule les estimations des caractéristiques probabilistes inconnues des classes, qui sont ensuite utilisées à la place des caractéristiques réelles inconnues de l'objet classé, dans l'algorithme optimal de classification construit pour une information a priori complète.

Considérons deux classes $S_1$ et $S_2$ caractérisées par des distributions normales à $N$ dimensions, de même matrice des covariances $\mathbf{M}$ et de vecteurs des moyennes $\mathbf{a}_1$ et $\mathbf{a}_2$ différents.

Si $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{M}$ sont donnés, la règle optimale (de Bayes) de classification du vecteur observé $\mathbf{x}$ de dimension $N$ prescrit la comparaison avec un seuil du logarithme du rapport de vraisemblance, c'est-à-dire de la statistique (voir tome II, page 76)

$$V = \left(\mathbf{x} - \frac{\mathbf{a}_1 + \mathbf{a}_2}{2}\right)' \mathbf{M}^{-1} (\mathbf{a}_2 - \mathbf{a}_1). \tag{4.1}$$

La statistique $V$ est répartie suivant une loi normale de paramètres (valeur moyenne et variance)

$$(-d_N^2/2, d_N^2), \quad \mathbf{x} \in S_1, \tag{4.2}$$

$$(d_N^2/2, d_N^2), \quad \mathbf{x} \in S_2, \tag{4.2'}$$

où

$$d_N^2 = (\mathbf{a}_2 - \mathbf{a}_1)' \mathbf{M}^{-1} (\mathbf{a}_2 - \mathbf{a}_1). \tag{4.3}$$

La grandeur $d_N^2$ peut être appelée carré de la « distance » entre les classes (à $N$ dimensions).

La probabilité totale d'erreur de classification, lorsqu'on utilise cette règle, est égale à (voir (1.29))

$$P_{\text{er}} = p_1\left[1 - F\left(\frac{C}{d_N} + \frac{d_N}{2}\right)\right] + p_2 F\left(\frac{C}{d_N} - \frac{d_N}{2}\right), \tag{4.4}$$

où $p_1$, $p_2$ sont les probabilités a propri respectives des classes; $F(x)$ est l'intégrale de Laplace; $C$ le seuil qui dépend du critère choisi.

Si les quantités $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{M}$ sont inconnues, on peut utiliser la procédure adaptative de classification du vecteur $\mathbf{x}$. L'apprentissage avec maître donne deux échantillons vectoriels classifiés indépendants: $\mathbf{x}_1^{(1)}, \ldots, \mathbf{x}_{n_1}^{(1)}$ de la première distribution et $\mathbf{x}_1^{(2)}, \ldots, \mathbf{x}_{n_2}^{(2)}$ de la seconde. Pour estimations des vecteurs inconnus $\mathbf{a}_1$ et $\mathbf{a}_2$ et de la matrice des covariances $\mathbf{M}$, on prend les estimations du maximum de vraisemblance (voir tome II, 2.7.2), soit:

$$\hat{\mathbf{a}}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} \mathbf{x}_i^{(j)}, \quad j = 1\,;2, \tag{4.5}$$

$$\hat{\mathbf{M}} = \frac{1}{n_1+n_2-2}\left[\sum_{i=1}^{n_1}(\mathbf{x}_i^{(1)} - \hat{\mathbf{a}}_1)(\mathbf{x}_i^{(1)} - \hat{\mathbf{a}}_1)' + \right.$$
$$\left. + \sum_{i=1}^{n_2}(\mathbf{x}_i^{(2)} - \hat{\mathbf{a}}_2)(\mathbf{x}_i^{(2)} - \hat{\mathbf{a}}_2)'\right]. \tag{4.6}$$

En substituant $\hat{\mathbf{a}}_1$, $\hat{\mathbf{a}}_2$, $\hat{\mathbf{M}}$ dans (4.1) au lieu de $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{M}$, on obtient la statistique de classification suivante:

$$\hat{V} = \left(\mathbf{x} - \frac{\hat{\mathbf{a}}_1 + \hat{\mathbf{a}}_2}{2}\right)' \hat{\mathbf{M}}^{-1}(\hat{\mathbf{a}}_2 - \hat{\mathbf{a}}_1). \tag{4.7}$$

Pour une dimension donnée $N$ et lorsque les tailles $n_1$ et $n_2$ des échantillons d'apprentissage augmentent indéfiniment, la statistique $\hat{V}$ converge en probabilité vers la statistique $V$.

Nous allons étudier en détail les propriétés de la statistique de classification (4.7) dans le cas où la taille des échantillons d'apprentissage est limitée.

**4.1.2. Cas unidimensionnel.** Commençons par le cas le plus simple de $N = 1$, quand les classes $S_1$ et $S_2$ sont caractérisées par des distributions unidimensionnelles de moyennes $a_1$ et $a_2$ inconnues et de variances données, égales à $\sigma^2$.

L'apprentissage avec maître a donné deux échantillons d'apprentissage indépendants classifiés: $x_1^{(1)}, \ldots, x_{n_1}^{(1)}$ de la classe $S_1$ et $x_1^{(2)}, \ldots, x_{n_2}^{(2)}$ de la classe $S_2$. Choisissons pour estimations des va-

leurs moyennes inconnues les estimations du maximum de vraisemblance qu'on trouve à partir des échantillons d'apprentissage (4.5) :

$$\hat{a}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i^{(1)}, \quad \hat{a}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} x_i^{(2)}. \tag{4.8}$$

En utilisant ces estimations au lieu des valeurs moyennes inconnues, on peut formuler la règle suivante de classification : l'observation $x$ appartient à la classe $S_2$ si

$$\hat{V} = \frac{1}{\sigma^2} \left( x - \frac{\hat{a}_1 + \hat{a}_2}{2} \right) (\hat{a}_2 - \hat{a}_1) \geqslant C \tag{4.9}$$

et à la classe $S_1$ dans le cas contraire.

Avec le critère du maximum de vraisemblance la formule (4.9) peut s'écrire sous forme d'un système d'inégalités :

$$x \geqslant \frac{\hat{a}_1 + \hat{a}_2}{2}, \quad \hat{a}_2 > \hat{a}_1, \tag{4.10}$$

$$x < \frac{\hat{a}_1 + \hat{a}_2}{2}, \quad \hat{a}_2 < \hat{a}_1, \tag{4.11}$$

i.e. pour le critère envisagé l'algorithme de classification consiste à comparer la valeur observée $x$ avec un seuil dépendant des échantillons d'apprentissage.

La statistique $\hat{V}$ dans (4.9) est le produit des variables aléatoires normales corrélées suivantes :

$$y = (\hat{a}_2 - \hat{a}_1)/\sigma, \tag{4.12}$$

$$z = \left( x - \frac{\hat{a}_1 + \hat{a}_2}{2} \right) \Big/ \sigma. \tag{4.13}$$

La moyenne et la variance de la variable aléatoire $y$ sont

$$m_1\{y\} = d_1 = (a_2 - a_1)/\sigma, \tag{4.14}$$

$$M_2\{y\} = 1/n_1 + 1/n_2, \tag{4.15}$$

tandis que les moyennes et les variances conditionnelles (par rapport à l'observation $x$) de la variable aléatoire $z$ et le coefficient de corrélation de $y$ et $z$ sont

$$m_1\{z \mid S_2\} = \frac{a_2 - a_1}{2\sigma} = -m_1\{z \mid S_1\}, \tag{4.16}$$

$$M_2\{z \mid S_2\} = M_2\{z \mid S_1\} = 1 + \frac{1}{4n_1} + \frac{1}{4n_2}, \tag{4.17}$$

$$R_{yz} = \frac{m_1\{yz\} - m_1\{y\}\, m_1\{z\}}{[M_2\{y\}\, M_2\{z\}]^{1/2}} =$$

$$= \frac{1}{2} \left( \frac{1}{n_1} - \frac{1}{n_2} \right) \left[ \left( 1 + \frac{1}{4n_1} + \frac{1}{4n_2} \right) \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \right]^{-1/2}. \tag{4.18}$$

Remarquons que pour $n_1 = n_2$ les variables $y$ et $z$ sont indépendantes. Les moyennes conditionnelles de la statistique $\hat{V}$ sont

$$m_1\{\hat{V} \mid S_2\} = \frac{1}{2}\left(\frac{1}{n_1} - \frac{1}{n_2}\right) + \frac{(a_2 - a_1)^2}{2\sigma^2}, \tag{4.19}$$

$$m_1\{\hat{V} \mid S_1\} = \frac{1}{2}\left(\frac{1}{n_1} - \frac{1}{n_2}\right) - \frac{(a_2 - a_1)^2}{2\sigma^2}. \tag{4.19'}$$

Si $n_1 = n_2 = n$ on a

$$m_1\{\hat{V} \mid S_2\} = -m_1\{\hat{V} \mid S_1\} = \frac{(a_2 - a_1)^2}{2\sigma^2}, \tag{4.20}$$

$$M_2\{\hat{V} \mid S_2\} = M_2\{\hat{V} \mid S_1\} = \frac{2}{n}\left(1 + \frac{1}{2n}\right) + \left(1 + \frac{1}{n}\right)\frac{(a_2 - a_1)^2}{\sigma^2}. \tag{4.20'}$$

A partir de (4.19) et (4.19'), et en vertu de (4.1), pour $N = 1$ on a

$$m_1\{\hat{V} \mid S_i\} = m_1\{V \mid S_i\} + \frac{1}{2}\left(\frac{1}{n_1} - \frac{1}{n_2}\right),$$
$$i = 1\,;2\,;$$

et pour $n_1 \to \infty$, $n_2 \to \infty$

$$m_1\{\hat{V} \mid S_i\} = m_1\{V \mid S_i\}.$$

Si $n_1 = n_2 = n$, pour $n$ quelconque on a

$$m_1\{\hat{V} \mid S_i\} = m_1\{V \mid S_i\},$$
$$M_2\{\hat{V} \mid S_i\} = \frac{2}{n}\left(1 + \frac{1}{2n}\right) + \left(1 + \frac{1}{n}\right) M_2\{V \mid S_i\},$$

et pour $n \to \infty$

$$M_2\{\hat{V} \mid S_i\} \to M_2\{V \mid S_i\}.$$

On tire de (4.19), (4.19') que les paramètres de la distribution assez compliquée de la statistique $\hat{V}$ sont les moyennes *inconnues* $a_1$ et $a_2$. D'où l'impossibilité de principe d'une analyse probabiliste des caractéristiques de l'algorithme de classification (4.9). On peut seulement, après avoir surmonté les difficultés liées à la détermination de la fonction de répartition de $\hat{V}$, trouver la probabilité d'erreur

$$P_{\text{er}}(a_1, a_2) = p_1 P\{\hat{V} \geqslant C \mid S_1\} + p_2 P\{\hat{V} < C \mid S_2\}, \tag{4.21}$$

puis substituer dans (4.21) au lieu de $a_1$, $a_2$ leurs estimations $\hat{a}_1$, $\hat{a}_2$ (4.8) et obtenir ainsi l'*estimation* $\hat{P}_{\text{er}}$ de la probabilité d'erreur.

Lorsque les tailles des échantillons d'apprentissage $n_1$ et $n_2$ augmentent indéfiniment, les estimations $\hat{a}_1$ et $\hat{a}_2$ convergent en probabilité vers les valeurs moyennes $a_1$ et $a_2$ respectivement (car la moyenne échantillonnée est une estimation consistante de la moyenne).

Comme de plus les limites en probabilité des sommes, des différences et des produits de variables aléatoires sont égales aux sommes, différences et produits des limites, on trouve à partir de (4.9) que lorsque les tailles des échantillons d'apprentissage augmentent indéfiniment, la règle de classification envisagée revient, en probabilité, à la règle de Bayes de vérification de l'hypothèse simple concernant la valeur moyenne d'une variable aléatoire normale (voir tome II, page 41), $\hat{P}_{er}$ convergeant en probabilité vers la valeur de $P_{er}$ donnée par (4.4) lorsque les quantités $a_1$ et $a_2$ sont connues a priori :

$$P_{er} = p_1\left[1 - F\left(\frac{C}{d_1} + \frac{d_1}{2}\right) + p_2 F\left(\frac{C}{d_1} - \frac{d_1}{2}\right)\right], \tag{4.22}$$

où

$$d_1^2 = (a_2 - a_1)^2/\sigma^2. \tag{4.23}$$

**4.1.3. Cas multidimensionnel (la matrice des covariances est donnée).** Considérons le cas multidimensionnel, lorsqu'il y a lieu de décider à laquelle des deux distributions normales à $N$ dimensions, de vecteurs des moyennes inconnus et de matrices de corrélation *données* $\mathbf{M}_1 = \mathbf{M}_2 = \mathbf{M}$, appartient l'échantillon observé $\mathbf{x}$. Supposons qu'après apprentissage avec maître on ait obtenu des échantillons d'apprentissage classifiés : $\mathbf{x}_1^{(1)}, \ldots, \mathbf{x}_{n_1}^{(1)}$ de la première distribution et $\mathbf{x}_1^{(2)}, \ldots, \mathbf{x}_{n_2}^{(1)}$ de la seconde. Chaque élément de ces échantillons est un vecteur à $N$ dimensions. On prend pour estimations des vecteurs des moyennes les estimations du maximum de vraisemblance (4.5)

$$\hat{\mathbf{a}}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} \mathbf{x}_i^{(1)}, \quad \hat{\mathbf{a}}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} \mathbf{x}_i^{(2)}. \tag{4.24}$$

L'algorithme de classification revient à comparer avec un seuil la statistique (4.7)

$$\hat{V} = \left(\mathbf{x} - \frac{\hat{\mathbf{a}}_1 + \hat{\mathbf{a}}_2}{2}\right)' \mathbf{M}^{-1}(\hat{\mathbf{a}}_2 - \hat{\mathbf{a}}_1). \tag{4.25}$$

Introduisons les désignations

$$\rho = (n_2 - n_1)\,[(n_2 + n_1)(n_1 + n_2 + 4n_1n_2)]^{-1/2}, \tag{4.26}$$

$$k = (n_2 - n_1)/(4n_1n_2). \tag{4.26'}$$

La statistique de classification pour le cas envisagé peut s'écrire sous la forme [51]

$$\hat{V}_i = \frac{k}{\rho}\,[(1+\rho)\,\chi^2(N, 2\lambda_1^{(i)}) - (1-\rho)\,\chi^2(N, 2\lambda_2^{(i)})], \tag{4.27}$$

où $\chi^2(N, 2\lambda_1^{(i)})$, $\chi^2(N, 2\lambda_2^{(i)})$ sont des variables aléatoires indépendantes réparties suivant la loi du $\chi^2$ non centrée à $N$ degrés de liberté

et de paramètres d'écart $2\lambda_j^{(i)}$ donnés par la formule *)

$$2\lambda_j^{(i)} = \frac{n_1 n_2}{4\,[1-(-1)^j \rho]}\,[(n_1+n_2)^{-1/2} - \\ -(-1)^{i+j}(n_1+n_2+4n_1n_2)^{-1/2}]^2\, d_N^2, \qquad (4.28)$$

où $d_N^2$ est donné par (4.3) et

$$i = 1 \text{ si } \mathbf{x} \in S_1, \qquad (4.29)$$
$$i = 2 \text{ si } \mathbf{x} \in S_2.$$

Notons que le paramètre d'écart dépend tant des tailles des échantillons d'apprentissage que de la matrice de correlation $\mathbf{M}$ (par la « distance » $d_N^2$).

Pour $n_1 \to \infty$ et $n_2 \to \infty$ la distribution de $\hat{V}$ tend vers une loi normale de paramètres $(d_N^2/2, d_N^2)$ si $\mathbf{x} \in S_2$ et de paramètres $(-d_N^2/2, d_N^2)$ si $\mathbf{x} \in S_1$.

**4.1.4. Cas multidimensionnel (la matrice des covariances est inconnue).** Dans le cas où l'on ne connaît ni les vecteurs des moyennes de deux répartitions normales, ni leurs matrices de corrélation $\mathbf{M}_1 = \mathbf{M}_2 = \mathbf{M}$, on utilise la statistique de classification (4.25), dans laquelle au lieu de $\mathbf{M}$ on utilise l'estimation $\hat{\mathbf{M}}$ obtenue d'après des échantillons d'apprentissage, c'est-à-dire (voir (4.6))

$$\hat{V} = \left(\mathbf{x} - \frac{\hat{\mathbf{a}}_1 + \hat{\mathbf{a}}_2}{2}\right)' \hat{\mathbf{M}}^{-1}(\hat{\mathbf{a}}_2 - \hat{\mathbf{a}}_1), \qquad (4.30)$$

où

$$\hat{\mathbf{M}} = \frac{1}{n_1+n_2-2}\left[\sum_{i=1}^{n_1}(\mathbf{x}_i^{(1)} - \hat{\mathbf{a}}_1)(\mathbf{x}_i^{(1)} - \hat{\mathbf{a}}_1)' + \sum_{i=1}^{n_2}(\mathbf{x}_i^{(2)} - \hat{\mathbf{a}}_2)(\mathbf{x}_i^{(2)} - \hat{\mathbf{a}}_2)'\right]$$

et $\hat{\mathbf{a}}_1$, $\hat{\mathbf{a}}_2$ sont données par (4.24).

Mais des difficultés apparaissent, ce qui est lié à l'impossibilité d'utiliser la statistique de classification (4.30) pour $N$ et des tailles des échantillons d'apprentissage $n_1$, $n_2$ quelconques, car pour $N \geqslant n_1 + n_2 - 2$ la matrice $\hat{\mathbf{M}}$ (voir (4.30')) se trouve être dégénérée, et il n'existe donc pas de matrice inverse $\hat{\mathbf{M}}^{-1}$ figurant dans l'expression (4.30). Laissant de côté la question des algorithmes optimaux de classification pour des matrices des covariances dégénérées, conservons l'algorithme de Bayes et étudions la statistique de classification (4.30) sous la condition

$$N < n = n_1 + n_2 - 2. \qquad (4.31)$$

*) Il est facile de voir que $\lambda_j^{(i)} > 0$.

Comme montré dans [16, 41, 53], lorsque cette condition est vérifiée, la statistique de classification (4.30) peut s'écrire comme suit :

$$\hat{V}_i = \frac{n}{\chi^2_{n-N+1}} \left( \mathrm{k}_1 b_{12}^{(i)} + \mathrm{k}_2 b_{11}^{(i)} - \mathrm{k}_1 t_{n-N+2} \sqrt{\frac{b_{11}^{(i)} b_{22}^{(i)} - b_{12}^{(i)}}{n-N+2}} \right), \tag{4.32}$$

où $t_{n-N+2}$, $\chi^2_{n-N+1}$ sont des variables aléatoires indépendantes, réparties respectivement suivant la loi de Student et la loi du $\chi^2$ centrée, et $b_{jk}^{(i)}$ les éléments d'une matrice aléatoire, répartie suivant la loi de Wishart non centrée à $N$ degrés de liberté et de paramètres d'écart $m_{jk}$, $j, k = 1 ; 2$, avec

$$m_{11} = n_1 n_2 (n_1 + n_2)^{-1} d_N^2 = m_{22} (n_1 + n_2 + 1) n_1 / n_2, \tag{4.33}$$

$$m_{12}^{(i)} = (-1)^i m_{11} [(n_1 + n_2 + 1) n_1 / n_2]^{-1/2}, \tag{4.34}$$

$$\mathrm{k}_1 = \left( \frac{n_1 + n_2 + 1}{n_1 n_2} \right)^{1/2}, \quad \mathrm{k}_2 = \frac{n_2 - n_1}{2 n_1 n_2}, \tag{4.35}$$

$d_N^2$ étant donné par (4.3) et $i = 1$ ou $i = 2$ conformément à (4.29).

La distribution non centrée de Wishart est si compliquée qu'elle ne peut être recommandée pour les applications pratiques (voir [2]). Cependant la représentation des statistiques de classification sous la forme (4.32) (ainsi que (4.27) lorsque $\mathbf{M}$ est donnée) sera utilisée ultérieurement pour l'étude de leurs propriétés asymptotiques, lorsque la dimension du vecteur des observations $\mathbf{x}$ augmente indéfiniment.

**4.1.5. Cas multidimensionnel (les matrices des covariances sont différentes et inconnues).** Supposons donnés les vecteurs des moyennes de deux distributions normales à $N$ dimensions des classes $S_1$ et $S_2$, soit $\mathbf{a}_1 = \mathbf{a}_2 = \mathbf{a}$. Les matrices de corrélation $\mathbf{M}_1$ et $\mathbf{M}_2$ de ces distributions sont *inconnues* et $\mathbf{M}_1 \neq \mathbf{M}_2$. A l'aide des échantillons classifiés d'apprentissage $\mathbf{x}_1^{(1)}, \ldots, \mathbf{x}_{n_1}^{(1)}$ de la première distribution et $\mathbf{x}_1^{(2)}, \ldots, \mathbf{x}_{n_2}^{(2)}$ de la seconde on peut écrire les estimations du maximum de vraisemblance des matrices inconnues $\mathbf{M}_1$ et $\mathbf{M}_2$ (voir tome II, page 147)

$$\hat{\mathbf{M}}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} (\mathbf{x}_i^{(1)} - \mathbf{a}) (\mathbf{x}_i^{(1)} - \mathbf{a})', \tag{4.36}$$

$$\hat{\mathbf{M}}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} (\mathbf{x}_i^{(2)} - \mathbf{a}) (\mathbf{x}_i^{(2)} - \mathbf{a})'. \tag{4.37}$$

Pour classifier l'observation $\mathbf{x}$ on peut faire usage de l'algorithme optimal de vérification des hypothèses sur la matrice de corrélation de la distribution normale, en remplaçant les matrices de corréla-

tion inconnues $\mathbf{M}_1$, $\mathbf{M}_2$ par leurs estimations (4.36) et (4.37). On obtient alors la règle de classification suivante : l'observation $\mathbf{x}$ appartient à la classe $S_2$ si

$$(\mathbf{x}-\mathbf{a})'(\hat{\mathbf{M}}_2^{-1}-\hat{\mathbf{M}}_1^{-1})(\mathbf{x}-\mathbf{a}) \geqslant 2\ln C + \ln\frac{\det\hat{\mathbf{M}}_1}{\det\hat{\mathbf{M}}_2}. \qquad (4.38)$$

En introduisant le changement de variable

$$\mathbf{Y}=\hat{\mathbf{f}}'(\mathbf{x}-\mathbf{a}), \quad \mathbf{Y}'=(y_1, \ldots, y_N),$$

où la matrice $\hat{\mathbf{f}}$ est donnée par la relation

$$\hat{\mathbf{M}}_2\hat{\mathbf{f}}=\hat{\mathbf{M}}_1\hat{\mathbf{f}}\hat{\Lambda},$$

$\hat{\Lambda}$ étant la matrice diagonale dont les éléments $\hat{\lambda}_1, \ldots, \hat{\lambda}_N$ sont les racines de l'équation $\det[\hat{\mathbf{M}}_2 - \hat{\lambda}\hat{\mathbf{M}}_1] = 0$, on peut écrire l'inégalité (4.38) comme suit :

$$\hat{V}=\sum_{i=1}^{N}\left(1-\frac{1}{\hat{\lambda}_i}\right)y_i^2 \geqslant 2\ln C + \sum_{i=1}^{N}\ln\hat{\lambda}_i. \qquad (4.39)$$

La distribution de la statistique $\hat{V}$ dans (4.39) est examinée dans [57], où l'on montre, en particulier, que pour $n_1 \to \infty$ et $n_2 \to \infty$ la règle (4.39) tend en probabilité vers la règle optimale de vérification des hypothèses sur la matrice de corrélation lorsque $\mathbf{M}_1$ et $\mathbf{M}_2$ sont connues. Si les vecteurs des moyennes de deux distributions normales sont égaux entre eux et ne sont pas donnés, on utilisera à la place de $\mathbf{a}$ son estimation obtenue à partir des échantillons d'apprentissage, soit

$$\hat{\mathbf{a}}=(n_1\hat{\mathbf{a}}_1+n_2\hat{\mathbf{a}}_2)/(n_1+n_2), \qquad (4.40)$$

où $\hat{\mathbf{a}}_1$ et $\hat{\mathbf{a}}_2$ sont données par (4.24).

**4.1.6. Classification à alternatives multiples pour des distributions quelconques.** Voyons comment on peut utiliser l'apprentissage avec maître dans le cas où l'on a $m$ classes $S_1, \ldots, S_m$ caractérisées par des distributions à $N$ dimensions (non obligatoirement normales) $w_N(\mathbf{x} \mid S_1), \ldots, w_N(\mathbf{x} \mid S_m)$. La forme de chacune des distributions ainsi que tous les paramètres (ou une partie) sont donnés. Si l'on a obtenu un échantillon classifié, c'est-à-dire $m$ échantillons de vecteurs à $N$ dimensions sachant que $\mathbf{x}_1^{(k)}, \ldots, \mathbf{x}_{n_k}^{(k)}$ appartient à la classe $S_k$, $k = 1, \ldots, m$, on peut d'apres les échantillons d'apprentissage former les estimations des paramètres inconnus des distributions, et en portant ces estimations dans $w_N(\mathbf{x} \mid S_k)$, trouver les estimations des fonctions de distribution $\hat{w}_N(\mathbf{x} \mid S_k) = g(\mathbf{x}, \mathbf{x}_1^{(k)}, \ldots, \mathbf{x}_{n_k}^{(k)}$. Puis, en utilisant comme critère de qualité

de la classification le minimum du risque moyen (pour des probabilités a priori données $p_k$, $k = 1, \ldots, m$, d'appartenance de l'échantillon observé à la distribution respective et connaissant les pertes $\Pi_{ij}$) on arrive à la règle suivante de classification, en remplaçant dans la règle de Bayes (voir tome II, page 70) les fonctions de distribution $w_N(\mathbf{x} \mid S_k)$ par leurs estimations: l'échantillon observé $\mathbf{x}$ appartient à la classe $S_k$ si

$$\sum_{i=1}^{m} (\Pi_{ij} - \Pi_{ik})\, p_i \hat{w}_N(\mathbf{x} \mid S_i) \geqslant 0, \; j = 1, \ldots, m. \tag{4.41}$$

Supposons, par exemple, que l'on ait $m$ classes caractérisées par des distributions normales multidimensionnelles à moyennes inconnues $\mathbf{a}_k$ et matrices des covariances $\mathbf{M}_1 = \ldots = \mathbf{M}_m = \mathbf{M}$ et que l'on dispose d'échantillons d'apprentissage de chacune des distributions. Si l'échantillon $\mathbf{x}_1^{(k)}, \ldots, \mathbf{x}_{n_k}^{(k)}$ appartient à une distribution normale de vecteur des moyennes $\mathbf{a}_k$, et de matrice des covariances $\mathbf{M}$, on peut former les estimations du maximum de vraisemblance des paramètres inconnus (voir (4.5) et (4.6)):

$$\hat{\mathbf{a}}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \mathbf{x}_i^{(k)}, \tag{4.42}$$

$$\hat{\mathbf{M}} = \sum_{k=1}^{m} \sum_{i=1}^{n_k} (\mathbf{x}_i^{(k)} - \hat{\mathbf{a}}_k)(\mathbf{x}_i^{(k)} = \hat{\mathbf{a}}_k)' \Big/ \sum_{i=1}^{m} (n_i - 1) \tag{4.43}$$

et trouver les estimations des distributions

$$\hat{w}_N(\mathbf{x} \mid S_k) = (2\pi)^{-N/2} (\det \hat{\mathbf{M}})^{-1/2} \exp\Big[ -\frac{1}{2} (\mathbf{x} - \hat{\mathbf{a}}_k)' \hat{\mathbf{M}}^{-1} (\mathbf{x} - \hat{\mathbf{a}}_k) \Big], \quad k = 1, \ldots, m. \tag{4.44}$$

En posant $\Pi_{ij} = \Pi$, $i \neq j$, $\Pi_{jj} = 0$ on déduit de (4.41) la règle de classification suivante: l'échantillon observé appartient à la classe $S_k$ si

$$\hat{V}_{kj} = \ln \frac{\hat{w}_N(\mathbf{x} \mid S_k)}{\hat{w}_N(\mathbf{x} \mid S_j)} = \left(\mathbf{x} - \frac{\hat{\mathbf{a}}_k + \hat{\mathbf{a}}_j}{2}\right)' \hat{\mathbf{M}}^{-1} (\hat{\mathbf{a}}_k - \hat{\mathbf{a}}_j) \geqslant \ln \frac{p_j}{p_k}, \quad j = 1, \ldots, m; \; j \neq k. \tag{4.45}$$

Lorsque tous les $n_k \to \infty$, la distribution simultanée des statistiques $\hat{V}_{kj}$ devient normale (voir [2]).

Si la forme de la fonction de distribution $w_N(\mathbf{x} \mid S_k)$ est inconnue, on peut utiliser des échantillons d'apprentissage pour l'estimation de ces fonctions par la méthode donnée dans l'annexe 1.

**4.1.7. Méthode géométrique.** Les règles de classification étudiées sont basées sur des distributions normales de paramètres inconnus et sont optimales même lorsque la distribution n'est pas normale, si pour critère de qualité on utilise la « distance » minimale du vecteur des observations au vecteur des échantillons d'apprentissage [25]. Nous appellerons *distance* du vecteur des observations $\mathbf{x}$ à l'échantillon d'apprentissage $\mathbf{x}_1^{(i)}, \ldots, \mathbf{x}_{n_i}^{(i)}$ la grandeur

$$\frac{1}{n_i} \sum_{k=1}^{n_i} |\mathbf{x} - \mathbf{x}_k^{(i)}|^2,$$

c'est-à-dire la somme des carrés des composantes des vecteurs $\mathbf{x} - \mathbf{x}_k^{(i)}$, $k = 1, \ldots, n_i$.

Pour rendre compacts des échantillons d'apprentissage d'une classe donnée, servons-nous d'une transformation linéaire. Il s'agit de trouver une transformation linéaire $\mathbf{A}_i$ conservant le volume (c'est-à-dire une transformation dont le jacobien est égal à l'unité) pour laquelle la grandeur

$$r_i = \frac{1}{n_i(n_i-1)} \sum_{k=1}^{n_i} \sum_{j=1}^{n_i} |\mathbf{A}_i(\mathbf{x}_k^{(i)} - \mathbf{x}_j^{(i)})|^2 \tag{4.46}$$

est minimale. Dans [25] il est montré que la rotation $\mathbf{C}_i$ avec transformation ultérieure de diagonalisation $\mathbf{D}$ est la transformation cherchée. Les colonnes de la matrice $\mathbf{C}_i$ sont les vecteurs propres de l'estimation $\hat{\mathbf{M}}_i$ de la matrice $\mathbf{M}_i$, les éléments de la matrice diagonale $\mathbf{D}$ étant

$$d_{kj}^{(i)} = \left[\prod_{l=1}^{N} \sigma_l^{(i)}\right]^{1/N} \delta_{kj}/\sigma_k^{(i)}, \tag{4.47}$$

où $\sigma_l^{(i)}$ est l'écart quadratique moyen des vecteurs $\mathbf{x}_k^{(i)}$ dans la direction du $l$-ième vecteur propre de la matrice $\mathbf{C}_i$ ; $N$ est la dimension des vecteurs, et $\delta_{kj} = 1$ pour $k = j$ ; $\delta_{kj} = 0$ pour $k \neq j$.

La règle de classification peut maintenant se formuler comme suit : 1) à partir des échantillons d'apprentissage et de l'observation $\mathbf{x}$ on calcule les grandeurs

$$\rho_i = \frac{1}{n_i} \sum_{k=1}^{n_i} |\mathbf{D}_i\mathbf{C}_i'(\mathbf{x} - \mathbf{x}_k^{(i)})|^2, \quad i = 1, \ldots, m; \tag{4.48}$$

2) on trouve la plus petite d'entre elles $\rho_j = \min_i \rho_i$ ; 3) on rapporte l'observation à la classe $j$.

Dans [25] il est montré également que la règle mentionnée est équivalente à la règle basée sur la détermination du logarithme du

rapport de vraisemblance, calculé en supposant que les distributions sont normales, ayant pour vecteurs des moyennes et matrices de corrélation les paramètres échantillonnés respectifs (les probabilités a priori étant égales entre elles : $p_i = p$ et $\Pi_{ij} \neq \Pi$, $i \neq j$, $\Pi_{jj} = 0$).

**4.1.8. Algorithme de classification avec auto-apprentissage.** Reprenons le problème formulé au début de 4.1.2 en supposant que l'échantillon d'apprentissage $x_1, \ldots, x_n$ ne soit pas classifié. Il est légitime de supposer que l'apparition des classes $S_1$ et $S_2$ dans chacune des observations a priori est équiprobable, alors on peut considérer chaque élément de l'échantillon d'apprentissage comme appartenant à une distribution commune bimodale

$$w_1(x \mid a_1, a_2) = \frac{1}{2\sigma\sqrt{2\pi}}\{\exp[-(x-a_1)^2/(2\delta^2)] + \\ + \exp[-(x-a_2)^2/(2\delta^2)]\}. \quad (4.49)$$

La valeur moyenne de la variable aléatoire suivant la distribution (4.49) est

$$a = (a_1 + a_2)/2; \quad (4.50)$$

$a_1$ et $a_2$ étant inconnues, $a$ est également inconnue.

La moyenne échantillonnée

$$\hat{a} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad (4.51)$$

obtenue à partir de l'échantillon d'apprentissage non classifié, est une estimation non biaisée de la valeur moyenne $a$ de la loi (4.49).

En utilisant l'estimation (4.51) au lieu de la moyenne inconnue $a$ et en adoptant le critère du minimum de la probabilité d'erreur de classification (maximum de vraisemblance), on peut formuler la règle de classification suivante. L'observation $x$ appartient à la classe $S_2$ si

$$x \geqslant \hat{a}, \quad (4.52)$$

et à la classe $S_1$ dans le cas contraire.

Lorsque $n$ augmente indéfiniment, la règle formulée tend asymptotiquement vers la règle de Bayes de vérification de l'hypothèse simple sur la valeur moyenne d'une variable aléatoire normale.

L'algorithme de classification peut être généralisé au cas multidimensionnel à condition de conserver la symétrie sphérique de la densité de probabilité. Le problème est de rapporter l'échantillon observé $\mathbf{x}$ à l'une des distributions normales à $N$ dimensions de vecteurs des moyennes inconnus $\mathbf{a}_1$ et $\mathbf{a}_2$ et de matrices de corrélation données $\mathbf{M}_1 = \mathbf{M}_2 = \sigma^2\mathbf{I}$, où $\mathbf{I}$ est la matrice unité. Dans ce cas la

distribution multidimensionnelle commune de deux classes est

$$w_N(\mathbf{x} \mid \mathbf{a}_1, \mathbf{a}_2) = \frac{1}{2(2\pi\sigma^2)^{N/2}} \left\{ \exp\left[\frac{-(\mathbf{x}-\mathbf{a}_1)'(\mathbf{x}-\mathbf{a}_1)}{2\sigma^2}\right] + \right.$$

$$\left. + \exp\left[\frac{-(\mathbf{x}-\mathbf{a})'(\mathbf{x}-\mathbf{a}_2)}{2\sigma^2}\right]\right\}; \quad (4.53)$$

$$w_N(\mathbf{x} \mid \mathbf{a}, \mathbf{b}) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[\frac{\mathbf{b}'\mathbf{b}}{2\sigma^2}\right] \exp\left[\frac{-(\mathbf{x}-\mathbf{a})'(\mathbf{x}-\mathbf{a})}{2\sigma^2}\right] \times$$

$$\times \operatorname{ch}\left[\frac{\mathbf{b}'(\mathbf{x}-\mathbf{a})}{\sigma^2}\right],$$

où

$$\mathbf{a} = \frac{\mathbf{a}_1 + \mathbf{a}_2}{2}, \qquad \mathbf{b} = \frac{\mathbf{a}_2 - \mathbf{a}_1}{2}. \quad (4.54)$$

Le vecteur **a** est le vecteur des valeurs moyennes de la distribution (4.53), les éléments de la matrice de corrélation **M** de cette distribution sont

$$M_{ij} = \int_{-\infty}^{\infty} \ldots \int (x_i - a_i)(x_j - a_j) \times$$

$$\times w_N(\mathbf{x} \mid \mathbf{a}, \mathbf{b})\, dx_1 \ldots dx_N = b_i b_j + \sigma^2 \delta_{ij}, \quad (4.55)$$

où $\delta_{ij}$ est le symbole de Kronecker ; $a_i$ et $b_i$ sont les composantes des vecteurs **a** et **b** respectivement.

Si les vecteurs des moyennes $\mathbf{a}_1$, $\mathbf{a}_2$ sont donnés, la partition optimale de Bayes de l'espace échantillonné est effectuée par un hyperplan perpendiculaire à la ligne joignant les points $\mathbf{x} = \mathbf{a}_1$ et $\mathbf{x} = \mathbf{a}_2$ et qui divise cette ligne en deux (voir tome II, problème 1.6). L'observation **x** se rapporte à l'une ou à l'autre classe, suivant le signe de la grandeur $\mathbf{b}'(\mathbf{x} - \mathbf{a})$ (comparer avec (4.25)).

Dans le cas où les vecteurs des moyennes sont inconnus pour les deux classes, il y a lieu de remplacer **a** et **b** par leurs estimations obtenues lors de l'apprentissage. Lors de l'auto-apprentissage d'après un échantillon non classifié $\mathbf{x}_1, \ldots, \mathbf{x}_n$, ces estimations sont obtenues à partir de la moyenne échantillonnée et de la matrice de corrélation échantillonnée. L'estimation du vecteur des moyennes est

$$\hat{\mathbf{a}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i, \quad (4.56)$$

et les estimations des composantes du vecteur **b** se trouvent à partir du système d'équations

$$\hat{M}_{ij} = b_i b_j + \sigma^2 \delta_{ij},$$

où $\hat{M}_{ij}$ est l'élément de la matrice de corrélation échantillonnée

$$\hat{\mathbf{M}} = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{x}_i - \hat{\mathbf{a}})(\mathbf{x}_i - \hat{\mathbf{a}})'. \tag{4.57}$$

Remarquons que le nombre d'équations dans ce système sera en général supérieur au nombre des inconnues. Pour résoudre ces équations on peut utiliser la méthode des moindres carrés.

Le calcul de l'estimation du vecteur **b** peut être simplifié si l'on met à profit le fait que ce vecteur est le plus grand des vecteurs propres de la matrice de corrélation de la distribution (4.54) ; en outre, pour le cas envisagé de la symétrie sphérique, tous les nombres caractéristiques de la matrice de corrélation, sauf le plus grand, sont égaux entre eux.

Le problème se trouve simplifié lorsque le vecteur moyenne pour l'une des classes, par exemple $\mathbf{a}_1$, est donné (comme dans le problème de la détection d'un signal noyé dans un bruit, lorsque $\mathbf{a}_1 = 0$). On a alors

$$\hat{\mathbf{b}} = \hat{\mathbf{a}} - \mathbf{a}_1 = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i - \mathbf{a}_1. \tag{4.58}$$

## 4.2. PROPRIÉTÉS ASYMPTOTIQUES DES STATISTIQUES DE CLASSIFICATION LORSQUE LE NOMBRE D'INDICES INFORMATIONNELS AUGMENTE

**4.2.1. Position du problème.** Nous allons reformuler le problème de la classification des observations en termes de la théorie de la reconnaissance des images. Au lieu des observations on utilise un ensemble de grandeurs scalaires qui sont des fonctions ou des fonctionnelles des grandeurs observées. Cet ensemble est appelé *indices informationnels de la classe* (*vecteur des indices*). Il est clair que celui qui a pour mission d'élaborer un système de reconnaissance essaie de résoudre le problème avec le plus petit nombre possible d'indices informationnels. A première vue, il peut sembler qu'en augmentant la dimension du vecteur des indices, on réduira la probabilité de classification erronée. Il est également intéressant de comparer l'influence sur la qualité de la classification de l'augmentation de la dimension du vecteur des indices et des tailles des échantillons d'apprentissage.

Supposons que les classes $S_1$ et $S_2$ soient caractérisées par des distributions normales $N$-dimensionnelles du vecteur des indices, de paramètres $(\mathbf{a}_1, \mathbf{M})$ et $(\mathbf{a}_2, \mathbf{M})$ respectivement qui sont a priori inconnus. On dispose d'échantillons d'apprentissage (classifiés) des indices pour chacune des classes, soit $(\mathbf{x}_1^{(1)}, \ldots, \mathbf{x}_{n_1}^{(1)}$ et $(\mathbf{x}_1^{(2)}, \ldots$

. . ., $\mathbf{x}_{n_2}^{(2)}$), qui servent à former, à l'aide de (4.5) et (4.6), les estimations des vecteurs des moyennes et de la matrice des covariances $\mathbf{M}$. A partir du vecteur des indices observé $\mathbf{x}$ on forme la statistique (4.7). Suivant que cette statistique dépasse ou non un seuil donné, on rapporte le vecteur des indices observé à l'une des classes $S_2$ ou $S_1$.

Il est intéressant d'étudier les propriétés asymptotiques de la statistique (4.7) ainsi que la probabilité d'erreurs de classification lorsque le nombre d'indices $N$ augmente indéfiniment. Certaines considérations qualitatives permettent de supposer que les régularités qui ont lieu pour des lois normales restent dans l'essentiel les mêmes pour d'autres distributions lorsque le nombre d'indices augmente.

**4.2.2. Cas de la symétrie sphérique des classes.** Nous allons commencer par le cas le plus simple, où l'on voit déjà apparaître certaines des lois générales. Supposons que les indices sont indépendants, c'est-à-dire que $\mathbf{M} = \sigma^2\mathbf{I}$, où $\mathbf{I}$ est la matrice unité, est que les tailles des échantillons d'apprentissage sont égales : $n_1 = n_2 = n$. Dans ce cas la statistique de classification est (4.7)

$$\hat{V} = \left(\mathbf{x} - \frac{\hat{\mathbf{a}}_2 + \hat{\mathbf{a}}_1}{2}\right)' (\hat{\mathbf{a}}_2 - \hat{\mathbf{a}}_1)/\sigma^2, \tag{4.59}$$

i.e. c'est la somme de $N$ variables aléatoires indépendantes où chacune des composantes est le produit de deux variables aléatoires normales indépendantes, et la distribution de $\hat{V}$ lorsque $\mathbf{x} \in S_1$ coïncide avec la distribution de $\hat{V}$ lorsque $\mathbf{x} \in S_2$. Les moyennes conditionnelles et les variances de la statistique $\hat{V}$ sont égales à (comparer (4.20) et (4.20′))

$$m_1\{\hat{V}|S_2\} = \frac{1}{2\sigma^2}\sum_{i=1}^{N}(a_{1i} - a_{2i})^2 = \frac{d^2{}_N}{2} = -m_1\{\hat{V}|S_1\}, \tag{4.60}$$

$$\mathbf{a}_1 = (a_{11}, \ldots, a_{1N}), \quad \mathbf{a}_2 = (a_{22}, \ldots, a_{2N}),$$

$$M_2\{\hat{V}|S_2\} = M_2\{\hat{V}|S_1\} = \frac{2N}{n}\left(1 + \frac{1}{2n}\right) + d_N^2\left(1 + \frac{1}{n}\right), \tag{4.60′}$$

avec

$$d_N^2 = \frac{1}{\sigma^2}\sum_{i=1}^{N}(a_{1i} - a_{2i})^2. \tag{4.60″}$$

Pour $n$ donné et $N \to \infty$, la statistique $\hat{V}$ dans (4.59) est asymptotiquement normale de paramètres (valeur moyenne et variance)

$$(-1)^i d_N^2/2, \quad \frac{2N}{n}\left(1 + \frac{1}{2n}\right) + \frac{d_N^2}{n} + d_N^2,$$

où $i = 1$, si $\mathbf{x} \in S_1$ ; $i = 2$ si $\mathbf{x} \in S_2$.

Par conséquent, la distribution asymptotique de la statistique de classification $\hat{V}$, pour une taille finie des échantillons d'apprentissage et lorsque la dimension du vecteur des indices augmente indéfiniment, diffère de la statistique $V$ correspondant au cas où l'on dispose d'une information a priori complète sur la distribution des classes, par une augmentation de la variance de la quantité (voir (4.2), (4.2'))

$$\Delta d_N^2 = \frac{2N}{n}\left(1+\frac{1}{2n}\right)+\frac{d_N^2}{n}.$$

Il est évident que pour $N \to \infty$ la « distance » $d_N^2$ entre les classes augmente également (voir (4.60")). C'est pourquoi la probabilité d'erreur de classification pour $N \to \infty$ tendra vers zéro si le rapport du module de la moyenne $|m_1\{\hat{V}\}|$ à la moyenne quadratique $[M_2\{\hat{V}\}]^{1/2}$ augmente indéfiniment.

Introduisons la fonction $A(N)$:

$$d_N^2 = \sqrt{N}\,A(N). \tag{4.61}$$

A partir de (4.60) et (4.60'), on obtient alors

$$\delta_N/2 = |m_1\{\hat{V}\}|/[M_2\{\hat{V}\}]^{1/2} = \frac{d_N^2}{2}\left[d_N^2\left(1+\frac{1}{n}\right)+\right.$$
$$\left.+\frac{2N}{n}\left(1+\frac{1}{2n}\right)\right]^{-1/2} =$$
$$= \frac{A(N)}{2}\left[\frac{A(N)}{\sqrt{N}}\left(1+\frac{1}{n}\right)+\frac{2}{n}\left(1+\frac{1}{2n}\right)\right]^{-1/2}. \tag{4.62}$$

Pour $N \to \infty$ on a la relation asymptotique suivante pour la probabilité d'erreur de classification (comparer avec (4.4)):

$$P_{\text{er}} = p_1\left[1-F\left(\frac{C}{\delta_N}+\frac{\delta_N}{2}\right)\right]+p_2F\left(\frac{C}{\delta_N}+\frac{\delta_N}{2}\right). \tag{4.63}$$

On voit que la probabilité d'erreur tend vers zéro pour $N \to \infty$ si $\delta_N \to \infty$, c'est-à-dire si

$$\lim_{N\to\infty} A(N) = \lim_{N\to\infty} d_N^2/\sqrt{N} = \infty. \tag{4.64}$$

Autrement dit, l'accumulation des indices tend à rendre l'erreur de classification aussi petite que l'on veut si le carré de la « distance » $d_N^2$ entre les distributions augmente comme $N^{1/2+\varepsilon}$, $\varepsilon > 0$. Dans le cas contraire, lorsque le nombre d'indices informationnels augmente indéfiniment, la probabilité d'erreur de classification reste positive.

**4.2.3. Classification dans le cas d'indices corrélés (la matrice des covariances est donnée).** Nous allons maintenant renoncer à l'hy-

pothèse de l'indépendance des indices et de l'égalité des tailles des échantillons d'apprentissage, en conservant l'hypothèse selon laquelle la matrice des covariances **M** commune aux deux classes est donnée. Dans ce cas la statistique de classification (4.25) peut s'écrire sous la forme d'une combinaison linéaire de deux variables aléatoires indépendantes suivant chacune une loi du $\chi^2$ non centrée (voir (4.27)).

Une variable aléatoire, répartie suivant une loi du $\chi^2$ non centrée à $N$ degrés de liberté et de paramètre d'écart $2\lambda > 0$, peut s'écrire comme la somme (voir également 3.1.11)

$$\chi^2(N,\ 2\lambda) = \sum_{i=1}^{N} \xi_i^2, \tag{4.65}$$

où $\xi_i$ sont des variables aléatoires normales indépendantes de variances unité, de moyennes nulles pour $i \geqslant 2$ et $m_1\{\xi_1\} = \sqrt{2\lambda}$. Donc la variable aléatoire $\chi^2(N,\ 2\lambda)$ est asymptotiquement normale (pour $N \to \infty$) de paramètres

$$m_1\{\chi^2(N,\ 2\lambda)\} = m_1\{\xi_1^2\} + \sum_{i=2}^{N} m_1\{\xi_i^2\} = 1 + 2\lambda + N - 1 = N + 2\lambda, \tag{4.66}$$

$$M_2\{\chi^2(N,\ 2\lambda)\} = M_2\{\xi_1^2\} + \sum_{i=2}^{n} M_2\{\xi_i^2\} = \\ = 2 + 8\lambda + 2(N-1) = 2N + 8\lambda. \tag{4.67}$$

En vertu de cette dernière affirmation, la statistique de classification (4.27), pour $N \to \infty$ et des tailles des échantillons d'apprentissage fixées d'avance, est *asymptotiquement normale* de paramètres

$$m_1\{\hat{V}_i\} = \frac{k}{\rho}\,[2\rho N + 2(\lambda_1^{(i)} - \lambda_2^{(i)}) + 2\rho(\lambda_1^{(i)} + \lambda_2^{(i)})] = \\ = \frac{n_2 - n_1}{2n_1 n_2} N + (-1)^i d_N^2/2, \quad i = 1,\ 2, \tag{4.68}$$

$$M_2\{\hat{V}_i\} = \frac{k^3}{\rho^2}\,[4N(1+\rho^2) + 8\lambda_1^{(i)}(1+\rho)^2 + 8\lambda_2^{(i)}(1-\rho)^2] = \\ = N\,[(n_1 + n_2 + 4n_1 n_2)(n_1 + n_2) + \\ + (n_2 - n_1)^2]/(2n_1 n_2)^2 + d_N^2\left(1 + \frac{1}{n_i}\right), \quad i = 1,\ 2. \tag{4.69}$$

Notons que pour $n_1 = n_2 = n$ les formules (4.68) et (4.69) coïncident avec (4.60) et (4.60′), c'est-à-dire, lorsque les tailles des échantillons d'apprentissage sont égales, la dépendance fonctionnelle de la distribution asymptotique de la statistique de classification par rapport à la taille des échantillons d'apprentissage et à la « distance » pour $N \to \infty$ et des indices corrélés, est la même que pour des indices indépendants.

En vertu de (4.68), dans le cas envisagé on a un écart égal à $\frac{n_2-n_1}{2n_1n_2}N$ par rapport à la valeur moyenne de la statistique de classification établie lorsque les paramètres des classes sont entièrement donnés. C'est pourquoi il y a lieu d'utiliser une statistique de classification sans biais, à savoir

$$\hat{V}_i^* = \hat{V}_i - \frac{n_2-n_1}{2n_1n_2}N. \tag{4.70}$$

Pour la statistique (4.70), le rapport du module de la moyenne à la moyenne quadratique est égal à (comparer avec (4.62))

$$\delta_N^i/2 = |m_1\{\hat{V}_i^*\}|/[M_2\{\hat{V}_i^*\}]^{1/2} =$$
$$= \frac{A(N)}{2}\left[\frac{A(N)}{\sqrt{N}}\left(1+\frac{1}{n_i}\right) + \{[(n_1+n_2+4n_1n_2)(n_1+n_2)+ \right.$$
$$\left. + (n_2-n_1)^2]/(2n_1n_2)^2\}^{-1/2}, \quad i=1;\ 2, \tag{4.71}$$

où $A(N) = d_N^2/\sqrt{N}$.

Ainsi dans le cas des indices corrélés (la matrice de corrélation étant donnée) pour des tailles finies des échantillons d'apprentissage, l'accumulation d'indices amène également à une erreur de classification aussi petite que l'on veut si le carré de la « distance » entre les classes augmente comme $N^{1/2+\varepsilon}$, $\varepsilon > 0$. Dans le cas contraire, la probabilité d'erreur pour $N \to \infty$ tend vers une constante positive.

**4.2.4. Classification dans le cas d'indices corrélés (la matrice des covariances est inconnue).** Considérons enfin le cas général, lorsque ni les vecteurs des moyennes, ni les matrices des covariances ne sont pas connus. Comme signalé dans 4.1.4, la statistique de classification dans ce cas, la condition (4.31) étant vérifiée, est une transformation fonctionnelle du type (4.32) des variables aléatoires indépendantes réparties suivant les lois de Student, du $\chi^2$ et la loi de Wishart non centrée.

Nous allons étudier les propriétés asymptotiques de la statistique (4.32) lorsque le nombre $N$ d'indices informationnels et les tailles des échantillons d'apprentissage augmentent indéfiniment. Supposons que ces nombres croissent à la même vitesse, c'est-à-dire

$$\lim_{\substack{N\to\infty \\ n_1\to\infty}} N/n_1 = r_1 < \infty, \tag{4.72}$$

$$\lim_{\substack{N\to\infty \\ n_2\to\infty}} N/n_2 = r_2 < \infty. \tag{4.73}$$

De plus on suppose que la croissance des $N$ et $n_1$, $n_2$ est telle que la condition (4.31) de non-dégénérescence de la matrice $\hat{\mathbf{M}}$ est toujours vérifiée.

En utilisant le fait que dans ces conditions les statistiques $\chi_{n-N+1}$, $t_{n-N+2}$, $b_{ij}$ entrant dans la statistique de classification (4.32) sont asymptotiquement normales, on peut montrer que cette statistique est, elle, *asymptotiquement normale* de paramètres

$$m_1\{\hat{V}_i\} = \frac{r_1+r_2}{r_1+r_2-r_1r_2}\left(\frac{r_1-r_2}{2}+\frac{(-1)^i d_N}{2}\right), \tag{4.74}$$

$$M_2\{\hat{V}_i\} = \left(\frac{r_1+r_2}{r_1+r_2-r_1r_2}\right)^3 (r_1+r_2+d_N^2), \quad i=1\,;\,2. \tag{4.75}$$

Il est facile de voir que pour $r_1 = r_2 = 0$, c'est-à-dire lorsque les tailles des échantillons d'apprentissage croissent plus rapidement que le nombre d'indices informationnels, les expressions (4.74) et (4.75) coïncident avec la moyenne et la variance de la statistique de classification $V$, lorsque les paramètres des classes classifiées sont entièrement connus (voir (4.2) et (4.2')). Ceci correspond évidemment à la tendance mentionnée de $\hat{V}$ vers $V$ lorsque la taille des échantillons d'apprentissage augmente indéfiniment.

Tout comme dans le cas précédent, il est désirable d'éliminer l'écart ; on y parvient en introduisant au lieu de $\hat{V}_i$ une statistique de classification sans biais (comparer avec (4.70))

$$\hat{V}_i^* = \left(1-\frac{r_1r_2}{r_1+r_2}\right)\hat{V}_i - \frac{r_1-r_2}{2} = \left(1-\frac{N}{n_1+n_2}\right)\hat{V}_i - \frac{n_2-n_1}{2n_1n_2}N, \quad i=1\,;\,2. \tag{4.76}$$

Pour la statistique (4.76) on a

$$\delta_N/2 = |\,m_1\{\hat{V}_i^*\}\,|/[M_2\{\hat{V}_i^*\}]^{1/2} = A(N)\left[\frac{r_1+r_2}{r_1+r_2-r_1r_2}\left(\frac{r_1+r_2}{N}+\frac{A(N)}{\sqrt{N}}\right)\right]^{-1/2}. \tag{4.77}$$

En vertu de la formule (4.77) la régularité évoquée ci-dessus se manifeste également dans le cas d'une matrice des covariances inconnue, si seulement la condition (4.31) est vérifiée, c'est-à-dire si le nombre d'indices informationnels ne croît pas plus vite que la taille totale des échantillons d'apprentissage. L'accumulation de ces indices donne alors une erreur de classification aussi petite que l'on veut si le carré de la « distance » entre les classes croît comme $N^{1/2+\varepsilon}$, $\varepsilon > 0$. Dans le cas contraire, même lorsque les tailles des échantillons d'apprentissage augmentent indéfiniment, la probabilité d'erreur pour $N \to \infty$ tend vers une constante positive.

## 4.3. PROCÉDURES D'APPRENTISSAGE DE BAYES

**4.3.1. Méthode des probabilités a posteriori.** Supposons que les paramètres $\boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_m$ dont dépendent les distributions $w_N(\mathbf{x} \mid \boldsymbol{\vartheta}_1, S_1), \ldots, w_N(\mathbf{x} \mid \boldsymbol{\vartheta}_m, S_m)$ des classes soient des vecteurs *aléatoires*

indépendants, de distributions a priori $w(\vartheta_k)$, $k = 1, \ldots, m$, qui rassemblent nos connaissances initiales sur les paramètres des distributions. Soit un ensemble d'échantillons d'apprentissage $\mathbf{x}_{ap} = (\mathbf{x}_1, \ldots, \mathbf{x}_m)$, où l'élément $\mathbf{x}_k$ est le vecteur ligne $\mathbf{x}_k = (x_1^{(k)}, \ldots, x_{n_k}^{(k)})$ des échantillons classifiés, appartenant à la classe $S_k$. On l'utilise pour la correction des distributions à priori $w(\vartheta_k)$, c'est-à-dire pour la détermination des distributions a posteriori $w(\vartheta_k \mid \mathbf{x}_k, S_k)$. Supposons que les observations ont fourni le vecteur $\mathbf{x}$. Le problème est de rapporter l'observation $\mathbf{x}$ à l'une des classes $S_1, \ldots, S_m$.

La formule de Bayes permet de trouver la probabilité a posteriori d'appartenance de l'observation à la classe $S_k$ pour $\mathbf{x}_{ap}$, $\mathbf{x}$ donnés :

$$P\{S_k \mid \mathbf{x}_{ap}, \mathbf{x}\} = \frac{p_k W(\mathbf{x}_{ap}, \mathbf{x} \mid S_k)}{\sum_{k=1}^{m} p_k W(\mathbf{x}_{ap}, \mathbf{x} \mid S_k)}, \tag{4.78}$$

où $p_k$ est la probabilité a priori d'appartenance de $\mathbf{x}$ à la classe $S_k$. Connaissant les grandeurs $P\{S_1 \mid \mathbf{x}_{ap}, \mathbf{x}\}, \ldots, P\{S_m \mid \mathbf{x}_{ap}, \mathbf{x}\}$, on rapporte l'observation $\mathbf{x}$ à la classe $S_j$ pour laquelle la probabilité a posteriori est maximale, c'est-à-dire

$$P\{S_j \mid \mathbf{x}_{ap}, \mathbf{x}\} > P\{S_k \mid \mathbf{x}_{ap}, \mathbf{x}\} \tag{4.79}$$

pour tous les $k \neq j$.

Comme

$$W(\mathbf{x}_{ap}, \mathbf{x} \mid S_k) = W(\mathbf{x} \mid \mathbf{x}_{ap}, S_k)\, W(\mathbf{x}_{ap}), \tag{4.80}$$

on déduit de (4.78) et (4.79)

$$p_j W(\mathbf{x} \mid \mathbf{x}_{ap}, S_k) > p_k W(\mathbf{x} \mid \mathbf{x}_{ap}, S_k) \tag{4.81}$$

pour tous les $k \neq j$.

Ainsi, l'algorithme de classification consiste à calculer les grandeurs $p_k W(\mathbf{x} \mid \mathbf{x}_{ap}, S_k)$, $k = 1, \ldots, m$, et à rapporter l'observation $\mathbf{x}$ à celle des classes $S_j$ à laquelle correspond la maximale des grandeurs trouvées. Si les probabilités a priori $p_k$ sont égales entre elles, la classification pour un ensemble donné $\mathbf{x}_{ap}$ revient à la détermination de la classe $S_j$ pour laquelle l'échantillon observé $\mathbf{x}$ maximise la fonction de vraisemblance $W(\mathbf{x} \mid \mathbf{x}_{ap}, S_k)$. Cette dernière peut être considérée comme l'estimation des distributions inconnues des classes pour un ensemble donné d'échantillons d'apprentissage $\mathbf{x}_{ap}$.

Calculons la fonction $W(\mathbf{x} \mid \mathbf{x}_{\text{ap}}, S_k)$ par la formule de la probabilité totale :

$$W(\mathbf{x} \mid \mathbf{x}_{\text{ap}}, S_k) = \int_{\Omega_k} W(\mathbf{x} \mid \mathbf{x}_{\text{ap}}, \boldsymbol{\vartheta}_k, S_k)\, W(\boldsymbol{\vartheta}_k \mid \mathbf{x}_{\text{ap}}, S_k) d\boldsymbol{\vartheta}_k =$$

$$= \int_{\Omega_k} W(\mathbf{x} \mid \boldsymbol{\vartheta}_k, S_k)\, W(\boldsymbol{\vartheta}_k \mid \mathbf{x}_k, S_k)\, d\boldsymbol{\vartheta}_k, \boldsymbol{\vartheta}_k \in \Omega_k, \quad (4.82)$$

car il est évident que $W(\boldsymbol{\vartheta}_k \mid \mathbf{x}_{\text{ap}}, S_k) = W(\boldsymbol{\vartheta}_k \mid \mathbf{x}_k, S_k)$, tandis que la fonction $W(\mathbf{x} \mid \mathbf{x}_{\text{ap}}, \boldsymbol{\vartheta}_k, S_k)$ ne dépend pas des échantillons d'apprentissage.

La densité de probabilité a posteriori du paramètre $\boldsymbol{\vartheta}_k$, pour un échantillon d'apprentissage donné, se calcule par la formule de Bayes, à savoir

$$W(\boldsymbol{\vartheta}_k \mid \mathbf{x}_k, S_k) = \frac{W(\boldsymbol{\vartheta}_k)\, W(\mathbf{x}_k \mid \boldsymbol{\vartheta}_k)\, S_k}{\int_{\Omega_k} w(\boldsymbol{\vartheta}_k)\, W(\mathbf{x}_k \mid \boldsymbol{\vartheta}_k, S_k)\, d\boldsymbol{\vartheta}_k}. \quad (4.83)$$

Dans la formule (4.83) le second facteur montre l'influence de l'apprentissage avec maître sur la distribution a priori des paramètres $\boldsymbol{\vartheta}_k$, $k = 1, \ldots, m$.

Notons que lorsque les paramètres $\boldsymbol{\vartheta}_k$ sont donnés et égaux à $\boldsymbol{\vartheta}_k^*$, leurs densités de probabilité conditionnelles sont des fonctions delta

$$W(\boldsymbol{\vartheta}_k \mid \mathbf{x}_k, S_k) = \delta(\boldsymbol{\vartheta}_k - \boldsymbol{\vartheta}_k^*)$$

et en vertu de (4.82) on a

$$W(\mathbf{x} \mid \mathbf{x}_{\text{ap}}, S_k) = W(\mathbf{x} \mid \boldsymbol{\vartheta}_k^*, S_k),$$

comme il se devait. Dans ce cas l'algorithme de classification coïncide avec l'algorithme optimal de Bayes de vérification des hypothèses à alternatives multiples (voir tome II, 1.4.8).

**4.3.2. Détection d'un signal aléatoire noyé dans un bruit normal (cas unidimensionnel).** Pour illustrer la méthode de classification exposée ci-dessus, nous allons considérer le problème de la détection d'un signal inconnu, noyé dans un bruit normal additif non corrélé, de moyenne nulle et de variance égale à $\sigma^2$. Le signal est une variable aléatoire normale $a$, indépendante du bruit, de paramètres $(a_0, \sigma_0^2)$. Le problème de la détection consiste dans ce cas de rapporter l'observation $x$ soit à la distribution du mélange additif du signal et du bruit $W_1(x \mid a_0, \sigma_0^2 + \sigma^2)$, soit à la distribution du bruit $W_1(x \mid 0, \sigma^2)$.

Si la distribution a priori du signal est donnée et n'a pas besoin d'être corrigée par apprentissage, c'est-à-dire si l'on connaît les

paramètres du signal $a_0$, $\sigma_0^2$, alors

$$W_1(x \mid a_0, \sigma_0^2+\sigma^2) = \frac{1}{\sqrt{2\pi(\sigma_0^2+\sigma^2)}} \exp\left[-\frac{(x-a_0)^2}{2(\sigma_0^2+\sigma^2)}\right] \quad (4.84)$$

et

$$W_1(x \mid 0, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}. \quad (4.84')$$

Le choix d'une décision (classification), suivant le critère du maximum de la probabilité a posteriori, est alors un cas particulier de la règle de Bayes de vérification des hypothèses (détection d'un signal). On décide que le signal est présent si

$$W_1(x \mid a_0, \sigma_0^2+\sigma^2) \geqslant W_1(x \mid \sigma^2), \quad (4.84'')$$

et que seul le bruit est présent si c'est l'inégalité inverse de (4.84″) qui est vérifiée. En vertu de (4.84) à (4.84″) l'algorithme de classification peut s'écrire comme suit: l'observation $X$ appartient au mélange du signal et du bruit si

$$a_0 X + \frac{1}{2}\frac{\sigma_0^2}{\sigma^2} X^2 \geqslant C, \quad (4.85)$$

et au bruit si

$$a_0 X + \frac{1}{2}\frac{\sigma_0^2}{\sigma^2} X^2 < C, \quad (4.85')$$

où

$$C = \frac{a_0^2}{2} + \frac{\sigma_0^2+\sigma^2}{2} \ln\left(1+\frac{\sigma_0^2}{\sigma^2}\right).$$

S'il y a lieu de classifier non pas une observation, mais un échantillon $(x_1, \ldots, x_N)$, en conservant les hypothèses de l'indépendance des valeurs du signal et du bruit, il faut remplacer dans (4.85) et (4.85′) $X$ et $X^2$ par les sommes des valeurs échantillonnées observées et les carrés de ces valeurs (voir tome II, page 65). Le dispositif optimal de détection du signal se compose de deux blocs: d'un filtre accordé et d'un récepteur énergétique. Les valeurs de sortie sont sommées, puis la somme est comparée à un seuil.

Considérons maintenant le cas où les données a priori concernant les paramètres de la distribution du signal *sont précisées au cours de l'apprentissage avec maître*. Supposons que l'on ait un échantillon d'apprentissage $x_1, \ldots, x_n$ dont on sait qu'il appartient au mélange du signal et du bruit. La distribution a posteriori du signal est donnée par la formule suivante (voir tome II, page 136):

$$W_1(a \mid x_1, \ldots, x_n) = \left(\frac{n}{2\pi\sigma^2}\right)^{1/2} \left(1+\frac{\sigma^2}{n\sigma_0^2}\right)^{1/2} \exp\left\{-\frac{n}{2\sigma^2}\left(1+\right.\right.$$

$$\left.\left.+\frac{\sigma^2}{n\sigma_0^2}\right)\left[a-\left(1+\frac{\sigma^2}{n\sigma_0^2}\right)^{-1}\left(\frac{1}{n}\sum_{k=1}^{n} x_k + \frac{a_0\sigma^2}{n\sigma_0^2}\right)\right]^2\right\}, \quad (4.86)$$

et, tout comme la distribution a priori, elle est normale. La différence est en ce que les paramètres $a_n$, $\sigma_n^2$ de la distribution a posteriori dépendent de l'échantillon d'apprentissage, à savoir

$$a_n = m_1\{a \mid x_1, \ldots, x_n\} = \left(1 + \frac{\sigma^2}{n\sigma_0^2}\right)^{-1} \left(\frac{1}{n}\sum_{k=1}^{n} x_k + \frac{a_0\sigma^2}{n\sigma_0^2}\right), \quad (4.87)$$

$$\sigma_n^2 = M_2\{a \mid x_1, \ldots, x_n\} = \sigma^2/(n + \sigma^2/\sigma_0^2). \quad (4.88)$$

Par conséquent la fonction de vraisemblance $W(X \mid x_1, \ldots, x_n)$ suit une loi normale de paramètres $a_n$ et $\sigma_n^2$, et la règle de classification de l'observation $X$ à l'aide de l'échantillon d'apprentissage $x_1, \ldots, x_n$ s'obtient à partir de (4.85) et (4.85′) en remplaçant $a_0$ par la valeur de la moyenne conditionnelle $a_n$ de (4.87), et $\sigma_0^2$ par la valeur de la variance conditionnelle $\sigma_n^2$ de (4.88). Autrement dit, la structure du dispositif optimal de détection reste inchangée, l'apprentissage n'intervenant que dans les valeurs des paramètres des blocs de ce dispositif.

Pour $n \to \infty$, en vertu de (4.87) et (4.88) on a

$$a_n \sim \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \sigma_n^2 \to 0$$

et par conséquent lorsque la taille de l'échantillon d'apprentissage augmente indéfiniment le dispositif optimal de détection tend à être linéaire (le second terme dans le membre de gauche de (4.85) tendra vers zéro). L'observation se rapporte au mélange du signal et du bruit si $\hat{a}X \geqslant \hat{a}^2/2$, où $\hat{a} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Cette règle coïncide évidemment avec (4.9) pour $a_2 = a$, $a_1 = 0$.

**4.3.3. Détection d'un signal aléatoire corrélé noyé dans un bruit corrélé.** L'exemple envisagé peut être généralisé au cas des valeurs corrélées du signal et d'un bruit corrélé, en conservant l'hypothèse de l'additivité du bruit et de son indépendance par rapport au signal. Supposons que le vecteur des valeurs moyennes du bruit soit nul, et que sa matrice de corrélation soit $\mathbf{M}$. La distribution a priori normale multidimensionnelle du signal est caractérisée par le vecteur des moyennes $\mathbf{a}_0$ et la matrice de corrélation $\mathbf{M}_0$. La distribution a priori du mélange additif du signal et du bruit indépendants est également normale, de vecteur des moyennes $\mathbf{a}_0$ et de matrice de corrélation $\mathbf{M} + \mathbf{M}_0$. On dispose d'un échantillon vectoriel d'apprentissage $\mathbf{x}_1, \ldots, \mathbf{x}_n$ appartenant au mélange du signal et du bruit. Après l'apprentissage, la distribution a posteriori du signal est également normale, seuls changent le vecteur des moyennes et la matrice de corré-

lation (conditionnels), qui sont donnés par les relations de récurrence, vraies pour le vecteur des moyennes $\mathbf{a}_n$ et la matrice de corrélation $\mathbf{M}_n$ conditionnels de la distribution normale [2]:

$$\mathbf{a}_n = \mathbf{M}(\mathbf{M}_{n-1}+\mathbf{M})^{-1}\mathbf{a}_{n-1} + \mathbf{M}_{n-1}(\mathbf{M}_{n-1}+\mathbf{M})^{-1}\mathbf{x}_n, \tag{4.89}$$

$$\mathbf{M}_n = \mathbf{M}(\mathbf{M}_{n-1}+\mathbf{M})^{-1}\mathbf{M}_{n-1}. \tag{4.90}$$

On en déduit

$$\mathbf{a}_n = \frac{\mathbf{M}}{n}\left(\mathbf{M}_0+\frac{\mathbf{M}}{n}\right)^{-1}\mathbf{a}_0 + \mathbf{M}_0\left(\mathbf{M}_0+\frac{\mathbf{M}}{n}\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i, \tag{4.91}$$

$$\mathbf{M}_n = \frac{\mathbf{M}}{n}\left(\mathbf{M}_0+\frac{\mathbf{M}}{n}\right)^{-1}\mathbf{M}_0. \tag{4.92}$$

Les formules (4.91) et (4.92) sont particulièrement simples dans le cas où $\mathbf{M} = \lambda\mathbf{M}_0$, $\lambda$ étant une constante positive. On a alors

$$\mathbf{a}_n = \frac{\lambda}{\lambda+n}\mathbf{a}_0 + \frac{n}{\lambda+n}\cdot\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i, \tag{4.93}$$

$$\mathbf{M}_n = \frac{1}{\lambda+n}\mathbf{M}. \tag{4.94}$$

Dans le cas unidimensionnel (4.93) devient (4.87), et (4.94) coïncide avec (4.88), $\lambda = \sigma^2/\sigma_0^2$.

La fonction de vraisemblance $W(\mathbf{x} \mid \mathbf{x}_{\mathrm{ap}})$ est, dans le cas général, une fonction de répartition normale multidimensionnelle, caractérisée par le vecteur des moyennes $\mathbf{a}_n$ et la matrice de corrélation $\mathbf{M}_n$ donnés par les formules (4.91) et (4.92). Il y a lieu de souligner que seules les moyennes conditionnelles dépendent des échantillons d'apprentissage (plus exactement, des moyennes échantillonnées), tandis que la matrice de corrélation conditionnelle ne dépend que de la taille $n$ de l'échantillon d'apprentissage.

**4.3.4. Méthode euristique.** Considérons maintenant le problème de la détection d'un signal inconnu noyé dans des bruits, lorsque l'on procède à un apprentissage sans maître, l'échantillon d'apprentissage n'étant pas classifié. Ceci signifie que l'on ne sait pas d'avance si la valeur échantillonnée $x_i$ appartient au mélange du signal et du bruit ou seulement au bruit. Il devient alors difficile de trouver la densité de probabilité a posteriori $W_1(a \mid x_{\mathrm{ap}})$ du paramètre $a$ pour l'échantillon non classifié $\mathbf{x}_{\mathrm{ap}} = (x_1, \ldots, x_n)$. Il semble que l'on peut palier à cette difficulté en examinant une à une toutes les suites possibles $x_1, \ldots, x_n$ où tout terme $x_i$ peut appartenir soit au mélange du signal et du bruit (état $S_1$), soit seulement au bruit (état $S_0$). Pour toute suite particulière $\Pi_n^{(k)}$ d'états lors de l'apprentissage, on a

$$\Pi_n^{(k)} = (S^{(1)}, \ldots, S^{(n)}),$$

où chacun des éléments $S^{(j)}$, $j = 1, \ldots, n$, peut être soit $S_1$, soit $S_0$; la densité de probabilité a posteriori $W_1(a \mid \mathbf{x}_{\text{ap}}, \Pi_n^{(k)})$ se calcule par la méthode examinée dans 4.3.2, tout comme dans le cas de l'apprentissage avec maître. Le nombre de suites différentes pour un échantillon d'apprentissage de dimension $n$ est égal à $2^n$. On a alors

$$W_1(a \mid \mathbf{x}_{\text{ap}}; \Pi_n^{(k)}) = \sum_{k=1}^{2^n} W_1(a \mid \mathbf{x}_{\text{ap}}, \Pi_n^{(k)})\, P\{\Pi_n^{(k)} \mid \mathbf{x}_{\text{ap}}\}, \qquad (4.95)$$

où $P\{\Pi_n^{(k)} \mid \mathbf{x}_{\text{ap}}\}$ est la probabilité pour que dans l'échantillon d'apprentissage soit réalisée la suite d'états $\Pi_n^{(k)}$. (Cette probabilité peut facilement être calculée en supposant, par exemple, que l'apparition ou la non-apparition du signal, lors de l'apprentissage, soient indépendantes, et connaissant la probabilité a priori d'apparition du signal dans l'une quelconque des observations.)

En vertu de (4.95) le dispositif optimal de détection pour l'auto-apprentissage à l'aide d'un échantillon de taille $n$ doit contenir $2^n$ dispositifs identiques qui sont les mêmes que dans le cas de l'apprentissage avec maître (filtre adapté et récepteur énergétique). Les tensions de sortie des dispositifs mentionnés sont multipliées par les coefficients de pondération $P\{\Pi_n^{(k)} \mid \mathbf{x}_{\text{ap}}\}$ et les produits partiels sont sommés. Pour des échantillons d'apprentissage de taille de l'ordre de quelques dizaines, il devient problématique de réaliser un dispositif utilisant l'algorithme d'essais. Il est evident que lorsque l'échantillon d'apprentissage est petit, les règles de classification ne sont pas très bonnes, mais au fur et à mesure de l'augmentation de $n$ leurs caractéristiques tendent vers celles des algorithmes à statistique donnée de reconnaissance des classes. Cependant, la complexité du dispositif optimal, lors de l'apprentissage sans maître, croît alors comme $2^n$.

Il est montré dans [61] que pour calculer la densité de probabilité a posteriori d'un paramètre inconnu d'une classe, on ne peut indiquer un algorithme de classification fini, indépendant de la taille de l'échantillon d'apprentissage, que si les observations d'apprentissage $x_1, \ldots, x_n$ se caractérisent par une statistique suffisante de dimension finie. Pour le problème de détection d'un signal inconnu envisagé ici la densité de probabilité de tout élément de l'échantillon d'apprentissage (sans maître) a pour expression (voir (4.49))

$$w_1(x_i \mid a) = \frac{p}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x_i-a)^2}{2\sigma^2}\right] + \frac{1-p}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x_i^2}{2\sigma^2}\right), \qquad (4.96)$$

où $p$ est la probabilité de présence du signal. Mais pour une densité de probabilité du type (4.96) il n'existe pas de statistique suffisante de dimension finie [12]. Par conséquent, la complexité d'un dispositif optimal de détection utilisant un apprentissage sans maître

croît énormément avec l'augmentation de la taille de l'échantillon d'apprentissage.

Voyons comment on peut simplifier la règle de classification lors de l'apprentissage sans maître, n'oubliant pas que cette simplification ne peut être réalisée qu'au prix d'une perte d'optimalité. Remarquant que le dispositif optimal de détection d'un signal inconnu utilisant une procédure d'apprentissage avec maître a une structure suffisamment simple, et qu'au contraire, le dispositif analogue pour l'auto-apprentissage est très compliqué, il paraît naturel de recourir à une procédure consistant à deviner la suite d'états ($S_1$ — mélange du signal et du bruit, $S_0$ — bruit pur) lorsque l'on passe d'un élément de l'échantillon d'apprentissage à l'autre, et d'utiliser cette procédure comme si elle était correcte (comme si elle était elle-même « maître ») *).

La structure du dispositif de détection devient alors identique à celle du dispositif optimal de détection avec maître, les résultats d'essais lors de l'auto-apprentissage sont utilisés pour corriger les paramètres de ce dispositif conformément aux formules données dans 4.3.2. Comme les erreurs sont inévitables, le dispositif de détection n'est plus optimal. Cependant, comme noté dans [61], pour un fonctionnement satisfaisant d'un tel dispositif de détection, il n'est pas nécessaire que la suite d'états adoptée soit rigoureusement exacte.

Il a été démontré que, lorsque les échantillons d'apprentissage sont de taille $n$ importante, il existe un sous-ensemble de l'ensemble $2^n$ de toutes les suites d'états possibles, ayant les deux propriétés suivantes: 1) la probabilité d'avoir dans ce sous-ensemble une suite d'états réelle est voisine de l'unité; 2) si la suite choisie est un élément de ce sous-ensemble, le dispositif de détection utilisant la méthode euristique tend pour $n \to \infty$ vers la méthode optimale. Malheureusement, à l'heure actuelle on n'a pas encore trouvé de méthodes efficaces permettant d'indiquer les sous-ensembles doués des propriétés mentionnées (voir également [56]).

Une variante de la méthode euristique consiste à supposer que la suite des décisions adoptées au cours de l'auto-apprentissage est une suite euristique [59]. Dans cette méthode, on attribue initialement aux paramètres du dispositif de détection des valeurs arbitraires. Chaque fois que le dispositif prend la décision sur la présence du signal, et que le signal soit réellement présent à cet instant, les paramètres du dispositif sont corrigés en conséquence.

**4.3.5. Algorithme adaptatif bayesien de Robbins.** Pour utiliser la méthode usuelle de Bayes il faut connaître la distribution a priori des paramètres inconnus. Dans 4.3.1, la distribution a priori initiale

*) Cette méthode est parfois appelée apprentissage avec maître réel, c'est-à-dire avec un maître pouvant se tromper (voir également [23]).

du paramètre de la classe était modifiée en cours d'apprentissage avec maître. Posons la question de savoir s'il est possible d'employer l'apprentissage avec maître pour la solution des problèmes de classification afin d'utiliser la méthode de Bayes sans hypothèses supplémentaires sur la distribution a priori initiale. On donne une réponse affirmative à cette question grâce au théorème démontré par S. N. Bernstein et R. Mises dans [19]. Le théorème se résume en ce qui suit. Soit $x_1$ une valeur échantillonnée d'une distribution à paramètre $\vartheta$ inconnu (aléatoire). La distribution a posteriori de ce paramètre est

$$W_1(\vartheta \mid x_1) = w_1(\vartheta) \frac{W_1(x_1 \mid \vartheta)}{\int_{-\infty}^{\infty} w_1(\vartheta) W_1(x_1 \mid \vartheta)\, d\vartheta}. \tag{4.97}$$

où $w_1(\vartheta)$ est la densité de probabilité a priori du paramètre $\vartheta$. Si l'on extrait la valeur échantillonnée $x_2$, indépendante de $x_1$, alors $W_1(\vartheta \mid x_1)$ peut être utilisée en tant que nouvelle distribution a priori pour le calcul de la nouvelle distribution a posteriori

$$W_1(\vartheta \mid x_1, x_2) = W_1(x_1 \mid \vartheta) \frac{W_1(x_2 \mid \vartheta)}{\int_{-\infty}^{\infty} W_1(\vartheta \mid x_1) W_1(x_2 \mid \vartheta)\, d\vartheta} =$$

$$= w_1(\vartheta) \frac{W_1(x_1 \mid \vartheta) W_1(x_2 \mid \vartheta)}{\int_{-\infty}^{\infty} w_1(\vartheta) W_1(x_1 \mid \vartheta) W_1(x_2 \mid \vartheta)\, d\vartheta}. \tag{4.98}$$

D'une manière analogue, pour un échantillon indépendant $x_1, \ldots, x_n$ de taille $n$, on a *)

$$W_1(\vartheta \mid x_1, \ldots, x_n) = w_1(\vartheta) \frac{W_n(x_1, \ldots, x_n \mid \vartheta)}{\int_{-\infty}^{\infty} w_1(\vartheta) W_n(x_1, \ldots, x_n \mid \vartheta)\, d\vartheta}, \tag{4.99}$$

où

$$W_n \mid x_1, \ldots, x_n \mid \vartheta) = \prod_{k=1}^{n} W_1(x_k \mid \vartheta).$$

Le théorème mentionné ci-dessus affirme que si la densité de probabilité a priori $w_1(\vartheta)$ du paramètre $\vartheta$ est continue, au fur et à mesure de l'augmentation de la taille de l'échantillon, la densité a posteriori $W_1(\vartheta \mid x_1, \ldots, x_n)$ cesse de dépendre de la distribution a priori. Donc pour $n$ suffisamment grand, il devient peu important quelle

*) Notons que (4.99) est un cas particulier de (4.83).

fonction continue $w_1(\vartheta)$ utiliser dans (4.99). Ce théorème limite a constitué, sans doute, le pivot de la *méthode adaptative de Bayes* *) de la classification, proposée par Robbins [24].

Soit $(x_1, \vartheta_1), \ldots, (x_n, \vartheta_n)$ une suite de couples indépendants de variables aléatoires, tous les $\vartheta_i$, $i = 1, \ldots, n$, ayant la même distribution a priori $w_1(\vartheta)$, et tous les $x_i$ ayant la distribution

$$W_1(x) = \int_{-\infty}^{\infty} W_1(x \mid \vartheta)\, w_1(\vartheta)\, d\vartheta.$$

Pour des $x_i$ discrets, on a

$$P\{x_i = x\} = \int_{-\infty}^{\infty} P\{x_i = x \mid \vartheta\}\, w_1(\vartheta)\, d\vartheta.$$

Si l'on prend la décision $\gamma$ que l'observation $X$ appartient à la distribution $W_1(x \mid \vartheta)$, des pertes apparaissent déterminées par la fonction non négative $\Pi(\gamma, \vartheta)$. Comme $w_1(\vartheta)$ est inconnue, on fonde le choix d'une décision sur l'échantillon d'apprentissage $x_1, \ldots, x_n$ (les variables $\vartheta_1, \ldots, \vartheta_n$ restant toujours inconnues). Il est désirable que cette règle se rapproche pour $n \to \infty$ de la règle de Bayes dans le cas où $w_1(\vartheta)$ est donnée. Une règle satisfaisant à cette condition est dite adaptative, asymptotiquement optimale. Nous allons nous limiter à l'exposé de cette méthode dans le cas particulier du problème de la théorie des décisions statistiques à deux alternatives, à distribution a priori du paramètre $\vartheta$ inconnue, mais donnant la statistique des observations (dans [24] on peut trouver la formulation générale du problème).

Introduisons la désignation

$$\Delta(x) = \int_{-\infty}^{\infty} [\Pi(\gamma_1, \vartheta) - \Pi(\gamma_0, \vartheta)]\, W_1(x \mid \vartheta)\, w_1(\vartheta)\, d\vartheta \qquad (4.100)$$

et supposons que la fonction $\Delta_n(X)$ de l'observation $X$, dont la forme dépend de l'échantillon d'apprentissage $x_1, \ldots, x_n$, converge en probabilité vers $\Delta(X)$ pour $n \to \infty$. Dans ce cas on aura la règle adaptative asymptotiquement optimale suivante: on prend la décision $\gamma_0$ si

$$\Delta_n(X) \geqslant 0, \qquad (4.101)$$

et la décision $\gamma_1$ si c'est l'inégalité inverse de (4.101) qui est vérifiée. (Pour les distributions discrètes, dans (4.100) il y a lieu de remplacer $W_1(x \mid \vartheta)$ par $P\{x_i = x \mid \vartheta\}$.)

**4.3.6. Exemple d'algorithme de Robbins.** Voyons un exemple illustrant la possibilité de trouver la suite de fonctions $\Delta_n(x)$ convergeant en probabilité vers $\Delta(x)$. On vérifie l'hypothèse unilaté-

*) On dit aussi *méthode empirique de Bayes*.

rale $H$ que le paramètre $\vartheta$ d'une distribution de Poisson d'une variable aléatoire $x$ discrète ne dépasse pas une valeur donnée $\vartheta_0$, c'est-à-dire que $\vartheta \leqslant \vartheta_0$. Soit $\gamma_0$ la décision consistant à adopter l'hypothèse $H$, et $\gamma_1$ la décision rejetant cette hypothèse. La fonction de pertes $\Pi(\gamma_i, \vartheta)$ sera donnée comme suit :

$$\Pi(\gamma_0, \vartheta) = \begin{cases} 0, & \vartheta < \vartheta_0, \\ \vartheta - \vartheta_0, & \vartheta \geqslant \vartheta_0, \end{cases} \tag{4.102}$$

$$\Pi(\gamma_1, \vartheta) = \begin{cases} \vartheta_0 - \vartheta, & \vartheta \leqslant \vartheta_0, \\ 0, & \vartheta > \vartheta_0. \end{cases} \tag{4.102'}$$

Comme pour la distribution de Poisson d'une variable aléatoire discrète on a

$$P\{x_i = x \mid \vartheta\} = \frac{\vartheta^x}{x!} e^{-\vartheta}, \quad \vartheta > 0, \quad x = 0, 1, 2, \ldots, \tag{4.103}$$

en substituant (4.103) dans (4.100) (au lieu de $W_1(x \mid \vartheta)$) on obtient compte tenu de (4.102), (4.102')

$$\Delta(x) = \int_0^\infty (\vartheta_0 - \vartheta) \frac{\vartheta^x}{x!} e^{-\vartheta} w_1(\vartheta)\, d\vartheta =$$

$$= \vartheta_0 F_1(x) - (x+1) F_1(x+1), \tag{4.104}$$

où

$$F_1(x) = \int_0^\infty \frac{\vartheta^x}{x!} e^{-\vartheta} w_1(\vartheta)\, d\vartheta = P\{x_i = x\}. \tag{4.105}$$

Introduisons un compteur des coïncidences, à savoir

$$\nu(x, y) = \begin{cases} 1, & x = y, \\ 0, & x \neq y, \end{cases} \tag{4.106}$$

et considérons la fonction suivante de $x$, dépendant de l'échantillon d'apprentissage $x_1, \ldots, x_n$ :

$$u_n(x) = \frac{1}{n} \sum_{i=1}^{n} \nu(x, x_i). \tag{4.107}$$

La somme dans (4.107) est égale au nombre des valeurs échantillonnées d'apprentissage qui sont rigoureusement égales à $x$ et, par conséquent, $u_n(x)$ est une distribution empirique de $x$ convergeant en probabilité pour $n \to \infty$ vers $F_1(x)$ (voir (4.105)). La fonction

$$\Delta_n(x) = \vartheta_0 u_n(x) - (x+1)\, u_n(x+1) \tag{4.108}$$

converge en probabilité vers $\Delta(x)$ donnée par (4.104). On en déduit la règle adaptative asymptotiquement optimale suivante : si l'on a un

échantillon d'apprentissage $x_1, \ldots, x_n$, on prend la décision $\gamma_0$ (l'hypothèse $H$ est vraie) si

$$\frac{\vartheta_0}{n} \sum_{i=1}^{n} \nu(X, x_i) - \frac{X+1}{n} \sum_{i=1}^{n} \nu(X+1, x_i) \geqslant 0, \qquad (4.109)$$

et la décision $\gamma_1$ (l'hypothèse $H$ est fausse) si on a l'inégalité inverse de (4.109).

**4.3.7. Recherche du minimum du risque moyen à partir d'échantillons d'apprentissage.** Le risque moyen dans le problème de la vérification des hypothèses à alternatives multiples s'exprime par

$$R = \sum_{k=0}^{m} \sum_{j=0}^{m} \Pi_{jk} \int_{\mathbf{X}} \Phi_k(\mathbf{x})\, p_j W(\mathbf{x} \mid H_j)\, d\mathbf{x}, \qquad (4.110)$$

où $\Phi_k(\mathbf{x})$ est la $k$-ième composante de la fonction de décision vectorielle $\Phi(\mathbf{x})$

$$\Phi_k(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \in \mathbf{X}_k, \\ 0, & \mathbf{x} \in \mathbf{X}_j, \quad j \neq k, \quad j = 0, \ldots, m, \end{cases} \qquad (4.111)$$

$\mathbf{X}_k$ est le domaine de l'espace d'échantillonnage $\mathbf{X}$ correspondant à la prise de la décision $\gamma_k$ (adoption de l'hypothèse $H_k$); $\Pi_{jk}$ sont les éléments de la matrice de pertes: $p_j$ les probabilités a priori des hypothèses; $W(\mathbf{x} \mid H_j)$ les densités a posteriori des valeurs échantillonnées.

Si toutes les grandeurs du membre de droite de (4.110) sont connues, la règle minimisant le risque moyen se formule comme suit: on adopte l'hypothèse $H_k$ si (voir 4.1.6)

$$\sum_{i=0}^{m} (\Pi_{ij} - \Pi_{ik}) \frac{p_i}{p_0} \frac{W(\mathbf{x} \mid H_i)}{W(\mathbf{x} \mid H_0)} \geqslant 0, \quad j \neq k, \quad j = 0, \ldots, m. \qquad (4.112)$$

Le système d'inégalités (4.112) donne le domaine $\mathbf{X}_k$ pour la fonction de décision optimale.

Dans les conditions d'incertitude a priori, lorsque les caractéristiques probabilistes $p_j$, $W(\mathbf{x} \mid H_j)$ sont inconnues, on peut approcher le membre de gauche de l'équation (4.112) de la surface de séparation par une certaine fonction connue, dépendant du paramètre vectoriel $\boldsymbol{\vartheta}$, que l'on peut estimer (sous certaines restrictions) à partir de la condition du minimum du risque moyen, par la *méthode d'approximation stochastique* (voir annexe 2). Nous allons montrer que le problème de l'estimation du paramètre minimisant le risque moyen est identique au problème de la recherche d'un extremum (minimum) de la *fonction de régression*.

Introduisons la désignation

$$P_{kj}(\vartheta) = \int_{\mathbf{x}} \Phi_k(\mathbf{x}, \vartheta)\, p_j W(\mathbf{x} \mid H_j)\, d\mathbf{x}. \tag{4.113}$$

On peut alors écrire (4.110) comme suit :

$$R(\vartheta) = \sum_{k=0}^{m} \sum_{j=0}^{m} \Pi_{jk} P_{kj}(\vartheta). \tag{4.114}$$

La formule (4.114) peut être interprétée comme la valeur moyenne de la variable aléatoire discrète $z$ prenant $m^2$ valeurs possibles $\Pi_{jk}$, avec les probabilités $P_{kj}(\vartheta)$ correspondantes. La variable $z$ prend effectivement la valeur $\Pi_{jk}$ lorsque l'échantillon ayant la distribution $W(\mathbf{x} \mid H_j)$ appartient au domaine $\tilde{\mathbf{X}}_k$ déterminé par l'approximation adoptée (4.112) de la règle de Bayes. Sous forme analytique cette assertion s'écrit comme

$$z = f(\mathbf{x}) = \Pi_{jk}, \tag{4.115}$$

si $\mathbf{x}$ est un échantillon de la distribution $W(\mathbf{x} \mid H_j)$ et $\Phi_k(\mathbf{x}, \vartheta) = 1$, $k \neq j$.

Ainsi

$$R(\vartheta) = m_1 \{f(\mathbf{x}) \mid \vartheta\}, \tag{4.116}$$

et le problème de la détermination du paramètre $\vartheta^*$ optimal revient ainsi à minimiser la fonction de régression (4.116) pour des distributions inconnues $p_j W(\mathbf{x} \mid H_j)$. Pour trouver l'estimation $\vartheta^*$ par la méthode d'approximation stochastique, on a besoin d'un échantillon d'apprentissage indépendant $\mathbf{x}$ classifié, c'est-à-dire qu'il faut procéder à un apprentissage avec maître permettant de dire à quelle distribution $p_j W(\mathbf{x} \mid H_j)$, $j \neq 0, \ldots, m$, appartient chaque valeur échantillonnée $x_k$, $k = 1, 2, \ldots$

**4.3.8. Algorithme récurrent de discrimination de deux signaux inconnus.** Pous illustrer la méthode générale exposée dans 4.3.7, nous allons envisager le problème de la discrimination de deux signaux $S_0$ et $S_1$ dont on ne connaît ni les probabilités a priori $p_0$, $p_1 = 1 - p_0$, ni les fonctions de répartition $W_1(x \mid S_0)$, $W_1(x \mid S_1)$. Le problème de classification le plus simple consiste dans ce cas à identifier l'observation, qui est une grandeur scalaire, avec la valeur du signal $S_0$ (hypothèse $H_0$) ou avec la valeur du signal $S_1$ (hypothèse $H_1$). Effectuons le choix optimal de décision en conformité avec l'algorithme approché linéaire dépendant du paramètre scalaire inconnu $\vartheta$. Si $x < \vartheta$, l'observation doit être rapportée au signal $S_0$, et si $x \geqslant \vartheta$ au signal $S_1$. Soit un échantillon d'apprentissage classifié $x_1, \ldots, x_{2n}$. A l'aide de cet échantillon d'apprentissage, il y a lieu

de trouver pour la règle de classification formulée ci-dessus l'estimation récurrente du seuil optimal $\vartheta^*$ qui minimise le risque moyen.

Pour le problème envisagé le risque moyen est (voir (4.110) pour $m = 1$):

$$R(\vartheta) = \Pi_{00}\, p_0 [1 - \alpha(\vartheta)] + \Pi_{01}\, p_0\, \alpha(\vartheta) + \Pi_{10}\, p_1\, \beta(\vartheta) + \\ + \Pi_{11}\, p_1 [1 - \beta(\vartheta)], \tag{4.117}$$

où $\alpha(\vartheta)$, $\beta(\vartheta)$ sont les probabilités conditionnelles d'erreurs de première et de seconde espèce, égales à

$$\alpha(\vartheta) = \int_{-\infty}^{\infty} \Phi(x, \vartheta)\, W_1(x \mid S_0)\, dx, \tag{4.118}$$

$$\beta(\vartheta) = \int_{-\infty}^{\infty} [1 - \Phi(x, \vartheta)]\, W_1(x \mid S_1)\, dx, \tag{4.119}$$

et

$$\Phi(x, \vartheta) = \begin{cases} 1, & x \geqslant \vartheta, \\ 0, & x < \vartheta. \end{cases} \tag{4.120}$$

En vertu de (4.117) le risque moyen est une fonction de régression pour la variable aléatoire discrète $z = f(x)$, avec les quatre valeurs possibles suivantes:

$\Pi_{00}$ lorsque $\Phi(x, \vartheta) = 0$, i.e. $x < \vartheta$, et $x$ est une valeur échantillonnée issue de la distribution du signal $S_0$;

$\Pi_{01}$ lorsque $\Phi(x, \vartheta) = 1$, i.e. $x \geqslant \vartheta$, et $x$ est une valeur échantillonnée issue de la distribution du signal $S_0$;

$\Pi_{10}$ lorsque $\Phi(x, \vartheta) = 0$, i.e. $x < \vartheta$, et $x$ est une valeur échantillonnée issue de la distribution du signal $S_1$;

$\Pi_{11}$ lorsque $\Phi(x, \vartheta) = 1$, i.e. $x \geqslant \vartheta$, et $x$ est une valeur échantillonnée issue de la distribution du signal $S_1$.

Nous allons nous limiter au cas où la fonction $R(\vartheta)$ a un minimum unique. Ce cas a lieu, par exemple, pour $\Pi_{00} = \Pi_{11} = 0$, $\Pi_{10} = \Pi_{01} = 1$. En vertu de (4.117) on a alors

$$R(\vartheta) = p_0 \alpha(\vartheta) + p_1 \beta(\vartheta), \tag{4.121}$$

ce qui coïncide avec la probabilité totale d'erreur.

La courbe de la fonction de risque (4.121) (fig. 4.1) a un minimum unique pour $\vartheta = \vartheta^*$. Les valeurs asymptotiques de cette fonction sont $R(-\infty) = p_1$ et $R(\infty) = p_0$.

Pour estimer le seuil optimal $\vartheta^*$ dans les conditions d'indétermination a priori on peut utiliser la procédure itérative de l'approximation stochastique (voir (10) dans l'annexe 2). Soient $\vartheta_1$ une constante quelconque, $\{c_k\}$, $\{a_k\}$ des suites de nombres positifs, satisfaisant à la condition (8) de l'annexe 2, et $x_1, \ldots, x_{2n}$ un

échantillon d'apprentissage classifié. Dans ce cas l'estimation récurrente du seuil $\vartheta^*$ est de la forme

$$\vartheta_{k+1} = \vartheta_k - a_k \frac{z_{2k} - z_{2k-1}}{c_k}, \quad k = 1, 2, \ldots, n, \tag{4.122}$$

où $z_{2k} = i$ lorsque $x_{2k}$ appartient à la distribution du signal $S_{1-i}$ et $\Phi(x_{2k}, \vartheta_k + c_k) = i$, $i = 0; 1$; $z_{2k-1} = i$ lorsque $x_{2k-1}$ appartient à la distribution du signal $S_i$ et $\Phi(x_{2k-1}, \vartheta_k - c_k) = i$, $i = 0; 1$.

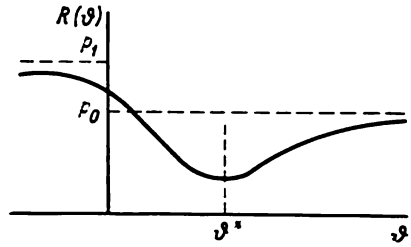



Fig. 4.1. Fonction de risque.

Lorsque la taille de l'échantillon d'apprentissage augmente indéfiniment ($n \to \infty$), l'estimation $\vartheta_n$ tend en probabilité vers le seuil $\vartheta^*$ minimisant le risque moyen (probabilité totale d'erreur) $R(\vartheta^*)$. Avec des restrictions supplémentaires à $R(\vartheta)$ on peut assurer la convergence en moyenne quadratique [43].

**4.3.9. Utilisation de la méthode des fonctions potentielles.** Considérons deux classes caractérisées par des distributions à $N$ dimensions $w_{N_1}(\mathbf{x})$ et $w_{N_2}(\mathbf{x})$, $\mathbf{x} \in \mathbf{X}$, et les probabilités a priori $p_1$, $p_2$ d'appartenance de $\mathbf{x}$ à la première ou la seconde classe respectivement. Si l'on connaît ces caractéristiques statistiques, le rapport de vraisemblance $l(\mathbf{x}) = p_2 w_{N_2}(\mathbf{x})/[p_1 w_{N_1}(\mathbf{x})]$ peut servir de statistique optimale de classification. Dans les conditions d'indétermination a priori, on peut essayer d'approcher (estimer) la statistique de classification à l'aide de l'échantillon d'apprentissage indépendant $\mathbf{y}_1, \ldots, \mathbf{y}_n$ de l'espace $\mathbf{X}$ *).

Ecrivons la statistique de classification inconnue $l(\mathbf{x})$ comme le développement suivant une base orthonormée finie $\{\varphi_1(\mathbf{x}), \ldots, \varphi_m(\mathbf{x})\}$ :

$$l(\mathbf{x}) \approx D(\mathbf{x}) = \sum_{k=1}^{m} a_k \varphi_k(\mathbf{x}). \tag{4.123}$$

*) Dans [34] on étudie le cas des échantillons d'apprentissage corrélés.

Introduisons la fonction potentielle (voir annexe 1)

$$K(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^{m} \varphi_k(\mathbf{x}) \varphi_k(\mathbf{y}); \tag{4.124}$$

il est montré dans [1] que $D(\mathbf{x})$ peut être approchée suffisamment bien par la fonction $D_n(\mathbf{x}, \mathbf{y}_n)$ obtenue par l'algorithme de récurrence suivant

$$D_n(\mathbf{x}, \mathbf{y}_n) = D_{n-1}(\mathbf{x}, \mathbf{y}_{n-1}) + r_{n-1} K(\mathbf{x}, \mathbf{y}_n), \tag{4.125}$$

où $\{r_k\}$ est une suite numérique satisfaisant à la condition de convergence de $D_n$ vers $D$ pour $n \to \infty$ (pour plus de détail voir [1]).

En désignant par $a_{kn}$ les coefficients du développement de $D_n(\mathbf{x}, \mathbf{y}_n)$ relativement à la base $\varphi(\mathbf{x})$ (voir (4.123)), on obtient à partir de (4.125) l'algorithme récurrent suivant d'estimation des coefficients $a_k$ du développement de la fonction $D(\mathbf{x})$ dans la base mentionnée :

$$a_{kn} = a_{k,\,n-1} + r_{n-1} \varphi_k(\mathbf{y}_n). \tag{4.126}$$

Pour commencer la procédure on prend pour le coefficient $a_{k0}$ une constante quelconque.

Dans [9] on examine un algorithme d'apprentissage, où l'on prend en qualité de fonction potentielle une fonction quelconque définie positive. La fonction $D(\mathbf{x})$ est un élément de l'espace hilbertien à noyau reproduisant généré par la fonction potentielle. Dans [9] on montre que les théorèmes de convergence démontrés pour l'algorithme (4.125) restent vrais pour une classe plus générale de fonctions potentielles.

**4.3.10. Approximation de la statistique de classification par la méthode de Parzen-Nadaraïa.** Soit $\mathbf{x}_1^{(i_1)}, \ldots, \mathbf{x}_n^{(i_n)}$ une suite d'échantillons d'apprentissage vectoriels (à $N$ dimensions) indépendants appartenant à la $i_j$- ième classe ($i_j = 1, 2$, $j = 1, \ldots, n$). Chacun des vecteurs $\mathbf{x}_j$ apparaît avec la probabilité $p_1$ si $i_j = 1$ et avec la probabilité $p_2$ si $i_j = 2$. Introduisons la variable aléatoire discrète $\nu_j$ pouvant prendre deux valeurs

$$\nu_j = \begin{cases} 1 & \text{si} \quad \mathbf{x}_j^{(i_j)} = \mathbf{x}_j^{(2)}, \\ 0 & \text{si} \quad \mathbf{x}_j^{(i_j)} = \mathbf{x}_j^{(1)}. \end{cases} \tag{4.127}$$

Il est évident que

$$P\{\nu_j = 1\} = p_2, \; P\{\nu_i = 0\} = p_1,$$
$$p_1 + p_2 = 1. \tag{4.128}$$

En introduisant la désignation

$$\mathbf{y}_j = \nu_j \mathbf{x}_j^{(1)} + (1 - \nu_j) \mathbf{x}_j^{(2)}, \tag{4.129}$$

on peut réunir les deux suites d'apprentissage en une seule $y_1^n = (y_1, \ldots, y_n)$ et utiliser cette suite pour approximation (estimation) de la statistique de classification $D(x)$.

Comme $D(x) \geqslant 0$, on peut employer pour estimer $D(x)$ les mêmes méthodes que dans le cas de l'estimation de la densité de probabilité multidimensionnelle (voir annexe 1, A. 1.2). Servons-nous pour approcher $D(x)$ de la fonction

$$D_n(x, y_1^n) = \frac{1}{n} \sum_{l=1}^{n} (2\nu_l - 1) K_l(x, y_l), \tag{4.130}$$

où

$$K_l(x, y_l) = \prod_{i=1}^{N} \left\{ K\left[\frac{x_i - y_{il}}{h_i(n)}\right] \Big/ h_i(n) \right\},$$

le noyau $K(z)$ et les constantes $h_i(n)$ vérifiant les conditions (18), (19), (22) et (22') de l'annexe 1, A. 1.2.

Il est montré dans [63] que la suite

$$\psi_n(y_1^n) = \int_X [D_n(x, y_1^n) - D(x)]^2 \, dx \tag{4.131}$$

converge en probabilité vers zéro lorsque la taille de l'échantillon d'apprentissage $n$ croît indéfiniment. Par ailleurs, l'estimation de la probabilité d'erreur de classification d'après la statistique de classification $D_n(x, y_1^n)$ converge vers la probabilité d'erreur de classification propre à la règle optimale utilisée avec une information a priori complète.

**4.3.11. Méthode adaptative d'élimination de l'indétermination a priori des paramètres perturbateurs.** Comme noté dans 1.1.7, dans certains problèmes pratiques de détection et de discrimination des signaux, les fonctions de vraisemblance des échantillons dépendent non seulement des paramètres informationnels, mais également des paramètres perturbateurs. Supposons que dans le problème de la vérification des hypothèses à alternatives multiples, la fonction de vraisemblance correspondant à la $j$-ième hypothèse, dépende du paramètre vectoriel perturbateur $\gamma_j$, $j = 0, 1, \ldots, m$. L'expression (4.110) du risque moyen sera dans ce cas une fonction de $\gamma_0^m = (\gamma_0, \ldots, \gamma_m)$, soit

$$R(\gamma_0^m) = \sum_{k=0}^{m} \sum_{j=0}^{m} \Pi_{jk} \int_X \Phi_k(x) \, p_j W(x \mid H_j, \gamma_j) \, dx. \tag{4.132}$$

Si l'on connaît les distributions a priori $W(\gamma_j)$ des paramètres perturbateurs, on peut (voir tome II, 1.4.2) exclure ces derniers par

moyennage sur $W(\gamma_j)$. On est amené alors au problème du choix optimal de décision, par minimisation du risque moyen, moyenné suivant les paramètres perturbateurs, i.e. il y a lieu de minimiser la quantité

$$R = \sum_{k=0}^{w} \sum_{j=0}^{w} \Pi_{jk} \int_{X} \Phi_k(\mathbf{x})\, p_j\, \widetilde{W}(\mathbf{x} \mid H_j)\, d\mathbf{x}, \tag{4.133}$$

où

$$\widetilde{W}(\mathbf{x} \mid H_j) = \int_{\Gamma} W(\mathbf{x} \mid H_j, \gamma_j)\, W(\gamma_j)\, d\gamma_j. \tag{4.134}$$

Dans les conditions d'indétermination a priori des paramètres perturbateurs, lorsque les distributions $W(\gamma_j)$ sont inconnues, on peut utiliser des échantillons d'apprentissage pour le calcul approché de l'intégrale (4.134). A cet effet, on trouve, à l'aide des échantillons d'apprentissage, les estimations du maximum de vraisemblance $\hat{\gamma}_j$ des paramètres perturbateurs. En supposant que les dérivées correspondantes existent, on a (voir [5])

$$\widetilde{W}(\mathbf{x} \mid H_j) \approx W(\mathbf{x} \mid H_j, \hat{\gamma}_j)\, W(\hat{\gamma}_j)\, (2\pi)^{N_j/2} [\det \mathbf{A}_j]^{-1/2}, \tag{4.135}$$

où $N_j$ est la dimension du vecteur $\gamma_j$ et $\mathbf{A}_j$ la matrice donnée par la formule

$$\mathbf{A}_j = -\left( \frac{\partial^2 \ln W(\mathbf{x} \mid H_j, \hat{\gamma}_j)}{\partial \gamma_{jk}\, \partial \gamma_{jl}} \right)_{\gamma_j = \hat{\gamma}_j}. \tag{4.136}$$

L'approximation (4.135) sera suffisamment bonne si les distributions a priori des paramètres perturbateurs sont bien plus larges que les distributions a posteriori établies d'après les échantillons d'apprentissage. La fonction $W(\hat{\gamma}_j)$ accuse une dépendance plus faible par rapport à $\mathbf{x}$ que les autres facteurs dans (4.135), ce qui permet de l'envisager, tout comme $(2\pi)^{N_j/2}$, comme un facteur négligeable. En vertu de (4.132) et (4.135), l'algorithme optimal de vérification des hypothèses est fondé sur la minimisation de la grandeur

$$\widetilde{R} = \sum_{k=0}^{m} \sum_{j=0}^{m} \Pi_{jk} \int_{X} \Phi_k(\mathbf{x})\, p_j\, W(\mathbf{x} \mid H_j, \hat{\gamma}_j)\, [\det \mathbf{A}_j]^{-1/2}\, d\mathbf{x}. \tag{4.137}$$

Pour une fonction de pertes simple

$$\Pi_{jk} = 1 - \delta_{jk}, \tag{4.138}$$

où $\delta_{jk}$ est le symbole de Kronecker, la règle optimale de choix d'une décision se formule comme suit : on adopte l'hypothèse $H_j$ si pour tous les $i \neq j$ on a

$$\Lambda_{ji}(\mathbf{x}) = \frac{p_j}{p_i} \frac{W(\mathbf{x} \mid H_j, \hat{\gamma}_j)}{W(\mathbf{x} \mid H_i, \hat{\gamma}_i)} \geqslant \left[ \frac{\det \mathbf{A}_j}{\det \mathbf{A}_i} \right]^{1/2}. \tag{4.139}$$

## 4.4. ALGORITHMES ADAPTATIFS ASYMPTOTIQUEMENT OPTIMAUX DE DÉTECTION DES SIGNAUX NOYÉS DANS DES BRUITS

**4.4.1. Principes de construction des algorithmes adaptatifs asymptotiquement optimaux.** Le critère général de qualité d'un algorithme adaptatif, comme nous l'avons noté dans 1.2.5, est sa consistance, c'est-à-dire sa convergence en probabilité vers l'algorithme optimal construit à partir de l'information a priori complète et lorsque la taille de l'échantillon d'apprentissage augmente indéfiniment.

On obtient des algorithmes adaptatifs consistants de détection à partir des algorithmes optimaux en utilisant au lieu des fonctions de distribution inconnues, ou de leurs paramètres, les estimations convenablement choisies et calculées à l'aide des échantillons d'apprentissage. Cependant, le critère de consistance ne donne pas d'une manière univoque l'algorithme adaptatif de détection de signaux.

On peut également déduire des algorithmes adaptatifs consistants à partir des algorithmes asymptotiquement optimaux de détection des signaux examinés au chapitre 3. Bien que ces algorithmes possèdent une certaine stabilité de structure, pour les utiliser on doit disposer d'une information a priori sur la distribution du bruit et son interaction avec le signal afin de trouver la caractéristique du convertisseur non linéaire (non inertiel pour un bruit indépendant ou inertiel pour un bruit corrélé) des observations et de calculer le seuil. A partir d'un échantillon d'apprentissage, on peut, en utilisant des estimations convenablement choisies de la caractéristique de non-linéarité mentionnée et du seuil, obtenir des algorithmes adaptatifs consistants de détection. Héritant toutes les qualités des algorithmes asymptotiquement optimaux, ils acquièrent une nouvelle propriété, à savoir la possibilité d'être utilisés en cas d'indétermination a priori quant au type du bruit.

Notons la particularité suivante des algorithmes adaptatifs asymptotiquement optimaux. Comme dans le cas des algorithmes asymptotiquement optimaux on suppose vérifiée la condition (1.50) de décroissance de l'amplitude du signal lors d'une augmentation illimitée de la taille $n$ de l'échantillon observé, l'estimation (suivant le coefficient d'efficacité asymptotique relative) d'après un échantillon d'apprentissage classifié du bruit est aussi efficace pour $n \to \infty$ que d'après un échantillon non classifié pouvant appartenir tant au bruit qu'au mélange du signal et du bruit *). Le fait de disposer d'un échantillon classifié (apprentissage avec maître) influe sur la qualité de l'algorithme utilisé dans le cas des $n$ finis. Autrement dit, la vitesse de convergence d'un algorithme adaptatif asymptotiquement

*) Ce qui s'explique par l'équivalence asymptotique signalée au chapitre 3 des mesures probabilistes lorsque l'hypothèse et l'alternative se rapprochent.

optimal, pour des estimations utilisant un échantillon classifié de bruit, sera plus grande qu'avec un échantillon non classifié.

Soient $\{\delta_n\}$ une suite d'algorithmes de détection de signaux, asymptotiquement optimaux pour une distribution donnée du bruit, et $\{\delta_{n,M}\}$ une suite d'algorithmes adaptatifs utilisant un échantillon classifié du bruit de taille fixée $M$. Désignons par $\rho_M(\delta, \delta_M)$ le coefficient d'efficacité asymptotique relative de la suite d'algorithmes adaptatifs $\{\delta_{n,M}\}$ par rapport aux algorithmes non adaptatifs $\{\delta_n\}$. Nous dirons que l'algorithme adaptatif $\delta_M$ est asymptotiquement optimal, si

$$\lim_{M\to\infty} \rho_M(\delta, \delta_M) = 1. \tag{4.140}$$

**4.4.2. Algorithme adaptatif asymptotiquement optimal de détection d'un signal déterministe noyé dans un bruit markovien $k$-fois connexe.** Soit $\mathbf{y}_1^M = (y_1, \ldots, y_M)$ un échantillon classifié appartenant à la distribution d'un bruit markovien $k$-fois connexe ($M > k$). Lorsque la distribution du bruit est inconnue, à partir de cet échantillon on peut obtenir une estimation des grandeurs suivantes, dépendant de la distribution inconnue, du développement asymptotique (3.211) :

$$\hat{f}_j(\mathbf{x}_{i-k}^i) = g_j(\mathbf{x}_{i-k}^i, \mathbf{y}_1^M), \tag{4.141}$$

$$\hat{K}_{lj} = \frac{1}{M}\sum_{i=1}^{M} \hat{f}_l(\mathbf{y}_{i-k}^i)\,\hat{f}_j(\mathbf{y}_{i-k}^i) =$$

$$= \frac{1}{M}\sum_{i=1}^{M} g_l(\mathbf{y}_{i-k}^i, \mathbf{y}_1^M)\, g_j(\mathbf{y}_{i-k}^i, \mathbf{y}_1^M). \tag{4.142}$$

La règle adaptative de détection d'un signal déterministe se formule comme suit : le signal est présent si

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \hat{\mathbf{f}}'(\mathbf{x}_{i-k}^i)\,\mathbf{s}_{i-k}^i - \frac{\gamma}{2}\operatorname{tr}[\mathbf{Q}\hat{\mathbf{K}}] \geqslant C, \tag{4.143}$$

et il n'y a pas de signal si c'est l'inégalité inverse de (4.143) qui est vérifiée.

Les composantes du vecteur des estimations $\hat{\mathbf{f}}$ et les éléments de la matrice des estimations $\hat{\mathbf{K}}$ sont donnés respectivement par (4.141) et (4.142).

On peut montrer que l'algorithme adaptatif de détection (4.143) est asymptotiquement optimal si pour toutes les valeurs $j = 0, \ldots, k$ on a

$$\lim_{M\to\infty} m_1\{[g_j(\mathbf{x}_{i-k}^i, \mathbf{y}_1^M) - f_j(\mathbf{x}_{i-k}^i)]^2\} = 0, \tag{4.144}$$

c'est-à-dire si les estimations $\tilde{f}_j$ convergent en moyenne quadratique vers $f_j$ lorsque la taille $M$ de l'échantillon d'apprentissage augmente indéfiniment. Dans (4.144), on prend la moyenne sur les produits des espaces des échantillons observés et des échantillons d'apprentissage pour l'hypothèse $H$. De plus, il faut que

$$\lim_{n\to\infty}\left|\frac{1}{\sqrt{n}}\sum_{i=0}^{n}\varphi^{l}_{i-k,i}\right|<\infty\,,\quad l=1,\ldots,m\,, \tag{4.145}$$

où $\varphi^{l}_{i-k,\,i}$ est le vecteur colonne de la matrice du signal $\varphi_{i-k,\,i}$ (voir (3.239)).

Au lieu de la condition (4.145) on peut exiger que les estimations $\hat{\mathbf{f}}\mathbf{x}^{i}_{i-k}$ soient *centrées*. Cependant, si l'espérance mathématique de l'estimation n'est pas donnée et il est impossible de la centrer, on devra se limiter aux signaux sans composante constante, comme il se devait conformément à (4.145).

**4.4.3. Algorithme adaptatif asymptotiquement optimal de détection d'un signal quasi déterministe noyé dans un bruit markovien $k$-fois connexe.** En utilisant (3.243), on peut formuler la règle adaptative de détection d'un signal quasi déterministe $\mathbf{s}=\boldsymbol{\vartheta}'\boldsymbol{\varphi}$: le signal est présent si

$$\int_{\Theta} w_m(\boldsymbol{\vartheta})\exp\left(\gamma\boldsymbol{\vartheta}'\hat{\mathbf{Y}}_n-\frac{\gamma^2}{2}\boldsymbol{\vartheta}'\,\hat{\mathbf{B}}\,\boldsymbol{\vartheta}\right)d\boldsymbol{\vartheta}\geqslant C, \tag{4.146}$$

et il n'y a pas de signal si c'est l'inégalité inverse de (4.146) qui est vérifiée. Dans l'inégalité (4.146) on a désigné

$$\hat{\mathbf{Y}}_n=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\mathbf{f}}'(\mathbf{x}^{i}_{i-k})\,\boldsymbol{\varphi}_{i-k,i} \tag{4.147}$$

et $\hat{\mathbf{B}}$ la matrice des estimations à éléments

$$\hat{\mathbf{B}}_{pq}=\operatorname{tr}[\hat{\mathbf{K}}\,\mathbf{A}_{pq}]. \tag{4.148}$$

$\hat{\mathbf{Y}}_n$ et $\hat{\mathbf{B}}$ s'obtiennent à partir de l'échantillon d'apprentissage du bruit $\mathbf{y}^{M}_{1}$.

Les composantes du vecteur des estimations $\hat{\mathbf{f}}$ et les éléments de la matrice des estimations $\hat{\mathbf{K}}$ sont donnés par (4.141) et (4.142). On peut montrer que les conditions (4.144) et (4.145) sont suffisantes pour que l'algorithme (4.146) de détection d'un signal quasi déterministe noyé dans un bruit markovien $k$-fois connexe soit asymptotiquement optimal.

Considérons maintenant quelques exemples de construction des algorithmes asymptotiquement optimaux de détection des signaux, utilisant les estimations satisfaisant à la condition (4.144).

**4.4.4. Algorithme adaptatif asymptotiquement optimal de détection d'un signal déterministe noyé dans un bruit additif à valeurs indépendantes (indétermination paramétrique).** Supposons que la densité de probabilité inconnue $w_1(x \mid 0)$ d'un bruit additif à valeurs indépendantes appartienne à une certaine famille paramétrique de densités. A titre d'exemple, on peut citer les fonctions de Pearson (voir par exemple [14]) ainsi que les densités susceptibles d'être approchées par des sommes finies de polynômes orthogonaux. Dans ces cas la dérivée logarithmique inconnue $f(x)$ de la densité de probabilité du bruit (voir (3.19)) peut s'écrire comme le rapport

$$f(x) = P(x)/Q(x), \tag{4.149}$$

où $P(x)$, $Q(x)$ sont des polynômes de degré fini dont les coefficients s'expriment en fonction des moments $m_k$, $k = 1, \ldots, N$, de la distribution $w_1(x \mid 0)$. En remplaçant ces moments a priori par les moments échantillonnés $m_k^*$, calculés d'après un échantillon d'apprentissage classifié du bruit :

$$m_k^* = \frac{1}{M} \sum_{i=1}^{M} y_i^k, \quad k = 1, \ldots, N,$$

on obtient l'estimation de la dérivée logarithmique :

$$\hat{f}(x) = P(x; m_1^*, \ldots, m_N^*)/Q(x; m_1^*, \ldots, m_N^*). \tag{4.150}$$

Cette estimation satisfait à la condition (4.144) qui dans le cas envisagé s'écrit

$$\lim_{M \to \infty} \int_{-\infty}^{\infty} [\hat{f}(x) - f(x)]^2 w_1(x \mid 0)\, dx = 0. \tag{4.151}$$

En plus de la correction de la caractéristique $f(x)$ du convertisseur non linéaire non inertiel d'entrée, il faut également corriger le seuil dont la grandeur dépend de la valeur inconnue $I_{f^2}$ (voir (3.13) et (3.15)).

Pour un échantillon d'apprentissage indépendant du bruit $y_i^M$, en vertu de la loi des grands nombres, l'estimation

$$\hat{I}_{f^2} = \frac{1}{M} \sum_{i=1}^{M} \hat{f}^2(y_i) \tag{4.152}$$

converge en probabilité vers $I_{f^2}$ lorsque la taille de l'échantillon d'apprentissage augmente indéfiniment.

**4.4.5. Algorithme adaptatif asymptotiquement optimal de détection d'un signal déterministe noyé dans un bruit additif à valeurs indépendantes (indétermination non paramétrique).** Supposons que

la densité de probabilité inconnue $w_1(x \mid 0)$ du bruit additif à valeurs indépendantes appartienne à une famille non paramétrique de densités. Pour construire un algorithme adaptatif asymptotiquement optimal de détection d'un signal, il y a lieu dans le cas envisagé de trouver l'estimation non pas de la densité de probabilité elle-même, mais de sa dérivée logarithmique. En utilisant la méthode des fonctions potentielles (voir annexe 1, A. 1.1.) on obtient l'estimation suivante de la caractéristique $f(x)$ du bloc non linéaire d'après l'échantillon d'apprentissage $\mathbf{y}_1^M$ du bruit :

$$\hat{f}(x) = \sum_{i=1}^{M} \sum_{k=1}^{m} \varphi_k'(x)\, \varphi_k(y_i) \Big/ \sum_{i=1}^{M} \sum_{k=1}^{m} \varphi_k(x)\, \varphi_k(y_i), \qquad (4.153)$$

où $\varphi_1(x), \ldots, \varphi_m(x)$ est une base orthonormale.

On obtient une autre estimation par la méthode de Parzen-Nadaraïa (voir annexe 1, A. 1.2) avec un noyau exponentiel :

$$\hat{f}(x) = \sum_{i=1}^{M} \frac{\partial}{\partial x} \exp\left[-\frac{(x-y_i)^2}{h(M)}\right] \Big/ \sum_{i=1}^{M} \exp\left[-\frac{(x-y_i)^2}{h(M)}\right], \qquad (4.154)$$

où $h(M) \to 0$ pour $M \to \infty$.

On peut montrer que les estimations (4.154) satisfont à la condition (4.151) et, par conséquent, les algorithmes adaptatifs utilisant ces estimations sont asymptotiquement optimaux.

**4.4.6. Algorithme adaptatif asymptotiquement optimal de détection d'un signal déterministe noyé dans un bruit additif markovien.** Supposons que la densité de probabilité de transition inconnue d'un bruit additif markovien (simplement connexe) appartienne à une famille exponentielle du type général [29,37] :

$$w(x_i \mid x_{i-1}) = \exp\Big\{ \Gamma(\alpha_1, \ldots, \alpha_N, \beta_{11}, \ldots, \beta_{NN}) + \\ + \sum_{j=1}^{N} \alpha_j \varphi_j(x_i) + \sum_{j=1}^{N} \sum_{l=1}^{N} \beta_{jl}\, \varphi_j(x_i)\, \varphi_l(x_{i-1}) \Big\}, \qquad (4.155)$$

$w$ dépend du paramètre vectoriel $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ et du paramètre matriciel $\boldsymbol{\beta} = (\beta_{jl})$ pour une famille donnée de fonctions linéairement indépendantes $\varphi_j(x)$, $j = 1, \ldots, N$. La fonction $\Gamma(\boldsymbol{\alpha}, \boldsymbol{\beta})$ se détermine à partir de la condition de normalisation

$$\int_{-\infty}^{\infty} w(x_i \mid x_{i-1})\, dx_i = 1.$$

Les dérivées logarithmiques de la densité de transition déterminant la statistique (3.223) sont dans ce cas

$$f_1(x_1, x_2) = \frac{\partial}{\partial s_i} \ln w(x_i - s_i \mid x_{i-1} - s_{i-1})\big|_{s_i = s_{i-1} = 0} =$$

$$= \sum_{j=1}^{N} \alpha_j \varphi_j'(x_i) + \sum_{j=1}^{N} \sum_{l=1}^{N} \beta_{jl} \varphi_j'(x_i) \varphi_l(x_{i-1}), \tag{4.156}$$

$$f_2(x_1, x_2) = \frac{\partial}{\partial s_{i-1}} \ln w(x_i - s_i \mid x_{i-1} - s_{i-1})\big|_{s_i = s_{i-1} = 0} =$$

$$= \sum_{j=1}^{N} \sum_{l=1}^{N} \beta_{jl} \varphi_j(x_i) \varphi_l'(x_{i-1}). \tag{4.157}$$

On peut montrer que les estimations $\hat{f}_1(x_1, x_2)$, $\hat{f}_2(x_1, x_2)$ obtenues à partir de (4.156) et (4.157) en substituant au lieu de $\alpha_j$ et $\beta_{jl}$ les estimations du maximum de vraisemblance, trouvées d'après un échantillon (d'apprentissage) classifié du bruit, satisfont à la condition (4.144). Pour la correction du seuil, il y a lieu de former également, toujours à l'aide de l'échantillon d'apprentissage, les estimations des grandeurs $K_{00}$, $K_{01}$, $K_{11}$ (voir (3.234)):

$$\hat{K}_{lj} = \frac{1}{M-1} \sum_{i=1}^{M-1} \hat{f}_l(y_i, y_{i+1}) \hat{f}_j(y_i, y_{i+1}), \quad l, j = 0; 1. \tag{4.158}$$

**4.4.7. Détection d'un signal quasi déterministe noyé dans un bruit additif markovien multiplement connexe, satisfaisant à l'équation d'autorégression.** Supposons que la suite des valeurs du bruit additif markovien satisfait à l'équation d'autorégression linéaire *) à coefficients inconnus

$$\sum_{j=0}^{k} \lambda_j x_{i-j} = \varepsilon_i, \quad i = k+1, k+2, \ldots, \tag{4.159}$$

où $\varepsilon_i$ est une suite de variables aléatoires indépendantes de densité de probabilité inconnue $w_\varepsilon(z)$. Dans ce cas la densité de probabilité de transition du bruit est

$$w(x_i \mid \mathbf{x}_{i-1}^{i-k}) = w_\varepsilon\left(\sum_{j=0}^{k} \lambda_j x_{i-j}\right),$$

*) Notons que le modèle envisagé dans 2.2.1 de l'autorégression est un cas particulier de (4.159) pour $k = 1$.

et la dérivée logarithmique de la densité de probabilité s'écrit

$$f_j(\mathbf{x}_{i-k}^i) = \lambda_j w_\varepsilon' \Big( \sum_{j=1}^{k} \lambda_j x_{i-j} \Big) \Big/ w_\varepsilon \Big( \sum_{j=1}^{k} \lambda_j x_{i-j} \Big), \quad j = 1, \ldots, k. \quad (4.160)$$

En portant (4.160) dans (3.215) on trouve la statistique vectorielle donnant l'algorithme asymptotiquement optimal de détection d'un signal quasi déterministe noyé dans un bruit additif satisfaisant à la condition (4.159) :

$$\mathbf{Y}_n(\mathbf{x}_{-k}^n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{R}(\varphi_{i-k,\,i})\, w_\varepsilon' [\mathbf{R}(\mathbf{x}_{i-k}^i)] / w_\varepsilon [\mathbf{R}(\mathbf{x}_{i-k}^i)]; \quad (4.161)$$

ou $\mathbf{R}$ désigne l'opérateur

$$\mathbf{R}(\mathbf{u}_{i-k}^i) = \sum_{j=1}^{k} \lambda_j u_{i-j}. \quad (4.162)$$

L'estimation $\hat{f}_j(x_{i-k})$ satisfaisant à la condition (4.144) est donnée par (4.160) à condition d'y remplacer les coefficients de régression inconnus $\lambda_j$ par leurs estimations du maximum de vraisemblance trouvées d'après un échantillon d'apprentissage du bruit.

Si $\{\varepsilon_i\}$ est une suite de variables aléatoires normales indépendantes, le système d'équations du maximum de vraisemblance par rapport aux coefficients de régression inconnus, où l'on a rejeté les termes convergeant vers zéro lorsque la taille $M$ de l'échantillon d'apprentissage augmente indéfiniment, est (voir [22, 34])

$$\sum_{j=0}^{k} \lambda_j r_j(\mathbf{y}) = \lambda_0^{-1},$$

$$\sum_{j=0}^{k} \lambda_j r_{|j-l|}(\mathbf{y}) = 0, \quad l = 1, 2, \ldots, k, \quad (4.163)$$

où $\mathbf{y} = (y_1, \ldots, y_M)$ est le vecteur des échantillons d'apprentissage et

$$r_j(\mathbf{y}) = \frac{1}{M} \sum_{i=j+1}^{M} y_i y_{i-j}. \quad (4.164)$$

Pour un bruit markovien biconnexe ($k = 2$) la solution du système (4.163) est

$$\hat{\lambda}_0 = \left( r_0 - r_1 \frac{r_0 r_1 - r_1 r_2}{r_0^2 - r_1^2} - r_2 \frac{r_0 r_2 - r_1^2}{r_0^2 - r_1^2} \right)^{-1/2}, \quad (4.165)$$

$$\hat{\lambda}_1 = \hat{\lambda}_0 \frac{r_1 r_2 - r_0 r_1}{r_0^2 - r_1^2}, \quad (4.165')$$

$$\hat{\lambda}_2 = \hat{\lambda}_0 \frac{r_1^2 - r_0 r_2}{r_0^2 - r_1^2}. \quad (4.165'')$$

## 4.5. PROPRIÉTÉS ASYMPTOTIQUES DES ESTIMATIONS DE BAYES DES PARAMÈTRES DU SIGNAL

**4.5.1. Position du problème.** Le modèle général du signal est donné par une fonction du type $s(t, \vartheta)$, où $\vartheta$ est un vecteur aléatoire appartenant au sous-ensemble $\Theta$ de l'espace euclidien, de densité de probabilité $w(\vartheta)$ donnée dans cet espace. Nous avons supposé que cette densité était connue à l'avance. En réalité il se peut que les paramètres aléatoires du signal ne soient pas connus, et dans ce cas le principe d'adaptation impose de substituer, dans les algorithmes optimaux, au lieu des paramètres inconnus des signaux, leurs estimations obtenues d'après des échantillons du signal mélangé avec du bruit (comme nous l'avons montré sur l'exemple de 4.3.2). Mais comment construire l'estimation du vecteur des paramètres du signal? Dans ce paragraphe nous allons montrer que pour une augmentation illimitée de la taille de l'échantillon d'apprentissage (du temps d'observation) les estimations de Bayes des paramètres du signal acquièrent des propriétés d'invariance par rapport aux variations de leurs distributions a priori et de leurs fonctions de pertes.

**4.5.2. Estimation bayesienne de l'amplitude d'un signal noyé dans un bruit normal additif pour un temps d'observation grand.** Commençons par un exemple simple où la fonction du signal $s(t, \vartheta) = as(t)$ dépend d'un seul paramètre aléatoire, à savoir de l'amplitude $\vartheta = a$. Pour une distribution a priori quelconque $w_1(a)$ de l'amplitude du signal, l'estimation de Bayes de cette amplitude, pour une fonction de pertes quadratique, est égale à (voir tome II, page 112)

$$\hat{a} = \int_{-\infty}^{\infty} a\, W[a/x(t)]\, da, \tag{4.166}$$

où $x(t)$ est la réalisation observée sur l'intervalle $(-T, T)$ du mélange additif du signal $as(t)$ et d'un bruit aléatoire normal de valeur moyenne nulle et de fonction de corrélation $B(t, y)$ donnée.

En portant dans (4.166) l'expression de la densité de probabilité a posteriori $W[a \mid x(t)]$ du paramètre estimé (voir tome II, 3.6.4), on trouve

$$\hat{a} = \frac{\int_{-\infty}^{\infty} a w_1(a) \exp\left(a x_T - \frac{a^2}{2} s_T\right) da}{\int_{-\infty}^{\infty} w_1(a) \exp\left(a x_T - \frac{a^2}{2} s_T\right) da}, \tag{4.167}$$

où

$$x_T = \int_{-T}^{T} V(t)\, x(t)\, dt, \quad s_T = \int_{-T}^{T} V(t)\, s(t)\, dt$$

et $V(t)$ est solution de l'équation intégrale

$$\int_{-T}^{T} B(t, u)\, V(u)\, du = s(t), \quad |t| \leqslant T.$$

En complétant les exposants des exponentielles dans (4.167) jusqu'à obtenir des carrés parfaits, on trouve après des transformations algébriques simples

$$\hat{a} = \frac{\int_{-\infty}^{\infty} a w_1(a) \sqrt{\frac{s_T}{2\pi}} \exp\left[-\frac{s_T}{2}\left(a - \frac{x_T}{s_T}\right)^2\right] da}{\int_{-\infty}^{\infty} w_1(a) \sqrt{\frac{s_T}{2\pi}} \exp\left[-\frac{s_T}{2}\left(a - \frac{x_T}{s_T}\right)^2\right] da}. \quad (4.168)$$

Comme pour $T \to \infty$ la grandeur $s_T$ croît indéfiniment, on a

$$\sqrt{\frac{s_T}{2\pi}} \exp\left[-\frac{s_T}{2}\left(a - \frac{x_T}{2}\right)^2\right] \sim \delta\left(a - \frac{x_T}{s_T}\right).$$

Puis en supposant $w_1(a)$ continue et en utilisant la propriété de filtration de la fonction delta, on obtient à partir de (4.168) une formule asymptotique pour l'estimation de Bayes de l'amplitude pour $T \to \infty$ :

$$\hat{a} \sim \frac{x_T}{s_T} = \int_{-T}^{T} V(t)\, x(t)\, dt \Big/ \int_{-T}^{T} V(t)\, s(t)\, dt. \quad (4.169)$$

Indépendamment de la forme de distribution a priori $w_1(a)$, l'estimation de Bayes $\hat{a}$ de l'amplitude $a$ converge pour $T \to \infty$ vers l'estimation du maximum de vraisemblance (voir tome II, 3.6.3).

Dans le cas où le processus aléatoire additif est un bruit blanc on a

$$s_T = \frac{1}{N_0} \int_{-T}^{T} s^2(t)\, dt,$$

c'est-à-dire que la convergence asymptotique signalée se trouve réalisée lorsque le rapport de l'énergie, sur l'intervalle d'observation, d'un processus déterministe à la densité spectrale du bruit blanc additif croît indéfiniment.

L'estimation sans biais du maximum de vraisemblance $\hat{a}_{mv} = x_T/s_T$ a une distribution normale de variance égale à $s_T^{-1}$. Par conséquent, pour $T \to \infty$ la densité de probabilité conditionnelle de l'estimation de Bayes peut s'écrire comme

$$W(\hat{a} \mid a) = \sqrt{\frac{s_T}{2\pi}} \exp\left[-\frac{s_T}{2}(\hat{a}-a)^2\right], \tag{4.170}$$

donc la densité de probabilité absolue $W_1(\hat{a})$ de l'estimation de Bayes $\hat{a}$ pour $T \to \infty$ tend vers $w_1(\hat{a})$:

$$W_1(\hat{a}) = \int_{-\infty}^{\infty} w_1(a)\, W(\hat{a} \mid a)\, da \sim w_1(\hat{a}),$$

i.e. vers la distribution a priori du paramètre inconnu, où la variable $a$ est remplacée par son estimation du maximum de vraisemblance.

**4.5.3. Propriétés asymptotiques de l'estimation bayesienne du paramètre vectoriel d'un signal noyé dans un bruit normal additif.** Nous allons envisager maintenant le cas plus général lorsque la réalisation observée est

$$x(t) = s(t, \boldsymbol{\vartheta}) + \xi(t), \tag{4.171}$$

où $s(t, \boldsymbol{\vartheta}) = \sum_{k=1}^{m} \vartheta_k \varphi_k(t) = \boldsymbol{\vartheta}'\boldsymbol{\varphi}(t)$, $\boldsymbol{\vartheta}$ un paramètre vectoriel dont la densité de probabilité est égale à $w_m(\boldsymbol{\vartheta})$; $\xi(t)$ un bruit aléatoire normal non stationnaire de valeur moyenne nulle et de fonction de corrélation donnée $B(t, y)$. Après discrétisation de la réalisation observée, on obtient un échantillon $\mathbf{x}$ de taille $n$

$$\mathbf{x} = \boldsymbol{\vartheta}'\boldsymbol{\varphi} + \mathbf{y},$$

où $\boldsymbol{\varphi}$ est une matrice $m \times n$ à éléments $\varphi_k(t_i)$, $k = 1, \ldots, m$, $i = 1, \ldots, n$, et $\mathbf{y}$ un échantillon vectoriel du bruit issu d'une loi normale de vecteur des moyennes nul et de matrice des covariances $\mathbf{B} = (B(t_i, t_j))$ de dimension $n \times n$. Dans la suite il sera commode d'utiliser la matrice de corrélation normée

$$\mathbf{b} = n\mathbf{B}/\operatorname{tr}\mathbf{B}, \tag{4.172}$$

ainsi que la grandeur

$$\nu = \boldsymbol{\vartheta}'\boldsymbol{\varphi}\boldsymbol{\varphi}'\boldsymbol{\vartheta}/\operatorname{tr}\mathbf{B}, \tag{4.173}$$

qui est le rapport signal/bruit généralisé.

La fonction de vraisemblance pour l'échantillon $\mathbf{x}$ (le vecteur $\boldsymbol{\vartheta}$ étant fixé) est égale à

$$w_n(\mathbf{x} \mid \boldsymbol{\vartheta}) = \left(2\pi \frac{\operatorname{tr}\mathbf{B}}{n}\right)^{-n/2} \times$$
$$\times \sqrt{\det \mathbf{b}} \exp\left\{-\frac{n}{2\operatorname{tr}\mathbf{B}}[(\mathbf{x}-\boldsymbol{\vartheta}'\boldsymbol{\varphi})'\mathbf{b}^{-1}(\mathbf{x}-\boldsymbol{\vartheta}'\boldsymbol{\varphi})]\right\}, \tag{4.174}$$

d'où l'on tire l'estimation du maximum de vraisemblance du vecteur des paramètres du signal (voir tome II, problème 3.6)

$$\hat{\vartheta}_{mv} = (\varphi' b^{-1} \varphi)^{-1} (\varphi' b^{-1} x). \tag{4.175}$$

L'estimation (4.175) est sans biais et sa matrice de corrélation est

$$m_1 \{(\hat{\vartheta}_{mv} - \vartheta)' (\hat{\vartheta}_{mv} - \vartheta)\} = (\varphi' b \varphi)^{-1} \frac{\operatorname{tr} B}{n}. \tag{4.176}$$

L'estimation de Bayes du vecteur $\vartheta$, pour une fonction de pertes quadratique, est comme on sait égale à (voir tome II, page 122)

$$\hat{\vartheta} = \int_{\Theta} \vartheta w_m(\vartheta) w_n(x \mid \vartheta) d\vartheta \Big/ \int_{\Theta} w_m(\vartheta) w_n(x \mid \vartheta) d\vartheta. \tag{4.177}$$

Il est facile de voir que la fonction de vraisemblance $w_n(x \mid \vartheta)$ peut s'écrire comme suit [17]:

$$w_n(x \mid \vartheta) = \exp\left\{-\frac{n}{2 \operatorname{tr} B} z' b^{-1} z\right\} \times$$
$$\times \exp\left\{-\frac{n}{2 \operatorname{tr} B} (\vartheta - \hat{\vartheta}_{mv})' D (\vartheta - \hat{\vartheta}_{mv})\right\}, \tag{4.178}$$

où

$$z = x - \varphi \hat{\vartheta}_{mv}, \tag{4.179}$$
$$D = \varphi' b^{-1} \varphi. \tag{4.180}$$

En substituant (4.178) dans (4.177) on obtient après des transformations évidentes

$$\hat{\vartheta} = \int_{\Theta} \vartheta w_m(\vartheta) \exp\left\{-\frac{n}{2 \operatorname{tr} B} (\vartheta - \hat{\vartheta}_{mv})' D (\vartheta - \hat{\vartheta}_{mv})\right\} d\vartheta \times$$
$$\times \left[\int_{\Theta} w_m(\vartheta) \exp\left\{-\frac{n}{2 \operatorname{tr} B} (\vartheta - \hat{\vartheta}_{mv})' D (\vartheta - \hat{\vartheta}_{mv})\right\} d\vartheta\right]^{-1},$$

et après le changement de la variable d'intégration $\vartheta - \hat{\vartheta}_{mv} = u$ on a

$$\hat{\vartheta} = \hat{\vartheta}_{mv} + \int_{U} u w_m(\hat{\vartheta}_{mv} + u) \exp\left(-\frac{n}{2 \operatorname{tr} B} u' D u\right) du \times$$
$$\times \left[\int_{U} w_m(\hat{\vartheta}_{mv} + u) \exp\left(-\frac{n}{2 \operatorname{tr} B} u' D u\right) du\right]^{-1}. \tag{4.181}$$

Nous nous limiterons à la classe de fonctions $w_m$ admettant le développement suivant:

$$w_m(\hat{\vartheta}_{mv} + u) = w_m(\hat{\vartheta}_{mv}) + u' w_m'(\hat{\vartheta}_{mv}) + \frac{1}{2} u' W u, \tag{4.182}$$

où $\mathbf{w}'_m(\hat{\boldsymbol{\vartheta}}_{mv})$ est le vecteur colonne à éléments

$$\frac{\partial}{\partial \vartheta_j} w_m(\boldsymbol{\vartheta}) \mid_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}}_{mv}}, \quad j=1, \ldots, m,$$

et $\mathbf{W}$ une matrice $m \times m$ à éléments

$$W_{ij} = \frac{\partial^2}{\partial \vartheta_i \, \partial \vartheta_j} w_m(\boldsymbol{\vartheta}^*),$$

$$\boldsymbol{\vartheta}^* = \boldsymbol{\Lambda}\boldsymbol{\vartheta} + (\mathbf{I} - \boldsymbol{\Lambda})\hat{\boldsymbol{\vartheta}}_{mv},$$

$\boldsymbol{\Lambda}$ étant une matrice diagonale à éléments $\lambda_k$ tels que $0 \leqslant \lambda_k \leqslant 1$, $k = 1, \ldots, m$, et $\mathbf{I}$ la matrice unité.

A l'aide de (4.182) on peut montrer que le second terme du membre de droite de (4.181) converge en probabilité vers zéro lorsque $n \to \infty$ (ou lorsque le rapport signal/bruit $\nu \to \infty$). Par conséquent, l'estimation du maximum de vraisemblance $\hat{\boldsymbol{\vartheta}}_{mv}$ est une estimation asymptotiquement optimale (pour $n \to \infty$) du paramètre vectoriel aléatoire du signal pour une fonction de pertes quadratique et pour toutes les fonctions $w_m(\boldsymbol{\vartheta})$ satisfaisant à la condition (4.182). Vu les propriétés bien connues des estimations de Bayes (voir tome II, page 114) on peut généraliser le résultat ci-dessus à d'autres fonctions de pertes symétriques (voir [18]).

Pour une réalisation continue de $x(t)$ observée sur un intervalle fini $(-T, T)$, l'estimation vectorielle du maximum de vraisemblance $\hat{\boldsymbol{\vartheta}}_{mv}$ dans (4.181) doit être calculée par la formule suivante (comparer avec (4.175), ainsi qu'avec 3.6.6 du tome II):

$$\hat{\boldsymbol{\vartheta}}_{mv} = \mathbf{s}_T^{-1}\mathbf{x}_T, \tag{4.183}$$

où $\mathbf{s}_T$ est une matrice à éléments

$$s_{T_{ij}} = \int_{-T}^{T} V_i(t)\,\varphi_j(t)\,dt, \tag{4.184}$$

$\mathbf{x}_T$ est le vecteur colonne à éléments

$$x_{T_i} = \int_{-T}^{T} V_i(t)\,x(t)\,dt \tag{4.185}$$

et $V_i(t)$ est solution de l'équation intégrale non homogène

$$\int_{-T}^{T} B(t, u)\,V_i(u)\,du = \varphi_i(t), \quad |t| \leqslant T. \tag{4.186}$$

**4.5.4. Propriétés asymptotiques de la distribution a posteriori du paramètre vectoriel d'un signal.** Considérons maintenant le cas général de l'estimation du paramètre vectoriel à $m$ dimensions $\boldsymbol{\vartheta}$

d'un signal, lorsque la distribution du bruit n'est pas obligatoirement normale. On sait que (voir tome II, chapitre 2), dans ce cas général, l'estimation du maximum de vraisemblance $\hat{\vartheta}_{mv}$, obtenue à partir d'un échantillon indépendant $(x_1, \ldots, x_n)$ de la distribution $w_n(x \mid \vartheta)$, est également asymptotiquement normale, de vecteur des moyennes égal à $\vartheta$ et de matrice de corrélation égale à la matrice inverse de la matrice informationnelle de Fisher. La fonction de vraisemblance de l'échantillon pour $n \to \infty$ peut être approchée par la densité de probabilité normale multidimensionnelle suivante (voir [31], § 12.7):

$$L_x(\vartheta) \sim (2\pi)^{-m/2} [\det Q(\hat{\vartheta}_{mv})]^{1/2} \times$$
$$\times \exp\left[-\frac{1}{2}(\vartheta - \hat{\vartheta}_{mv})' Q(\hat{\vartheta}_{mv})(\vartheta - \hat{\vartheta}_{mv})\right], \quad (4.187)$$

où $Q(\hat{\vartheta}_{mv})$ est une matrice à éléments

$$Q_{ij} = -\frac{\partial^2}{\partial\vartheta_i\,\partial\vartheta_j} \ln \prod_{k=1}^{n} w(x_k ; \vartheta)\,|_{\vartheta=\hat{\vartheta}_{mv}}. \quad (4.188)$$

Pour $n \to \infty$ la grandeur $\frac{1}{n} Q(\hat{\vartheta}_{mv})$ converge en probabilité vers la matrice informationnelle de Fisher $I(\vartheta)$, dont les éléments sont

$$I_{ij} = -m_1 \left\{\frac{\partial}{\partial\vartheta_i\,\partial\vartheta_j} \ln w(x ; \vartheta)\right\}. \quad (4.189)$$

Si la densité de probabilité a priori du paramètre $\vartheta$ est égale à $w_m(\vartheta)$, en vertu de la formule de Bayes la densité de probabilité a posteriori du paramètre $\vartheta$, calculée d'après l'échantillon $(x_1, \ldots, x_n)$ de la distribution $w(x ; \vartheta)$, est égale à

$$W(\vartheta \mid x_1^n) = \frac{w_m(\vartheta)\, L_x(\vartheta)}{\int_{\Theta} w_m(\vartheta)\, L_x(\vartheta)\, d\vartheta}. \quad (4.190)$$

Soit $\vartheta_0$ la valeur réelle du paramètre inconnu $\vartheta$ et supposons que $w_m(\vartheta_0) \neq 0$ et que $w_m(\vartheta)$ soit continue au voisinage de $\vartheta = \vartheta_0$. Pour un échantillon de grande taille, les variations de la densité a priori du paramètre $\vartheta$ au voisinage du point $\hat{\vartheta}_{mv}$ sont insignifiantes par rapport à la « pointe » de la fonction de vraisemblance $L_x(\hat{\vartheta}_{mv})$. C'est pourquoi, en vertu de (4.190), la densité de probabilité a posteriori du paramètre pour un échantillon de taille importante doit être assez bien approchée par la fonction de vraisemblance $W(\vartheta \mid x_1^n) \sim L_x(\vartheta)$ ou (compte tenu de (4.187)) par la densité de la distribution multidimensionnelle normale de vecteur des moyennes égal à $\hat{\vartheta}_{mv}$ et de matrice de corrélation $I^{-1}(\hat{\vartheta}_{mv})/n$.

En introduisant la normalisation

$$\bar{\vartheta} = (\vartheta - \hat{\vartheta}_{mv}) \sqrt{n}, \tag{4.191}$$

on obtient

$$\lim_{n \to \infty} w(\bar{\vartheta} \mid x_1^n) = (2\pi)^{-m/2} [\det \mathbf{I}(\vartheta_0)]^{1/2} \exp\left[-\frac{1}{2} \bar{\vartheta}' \mathbf{I}(\vartheta_0) \bar{\vartheta}\right]. \tag{4.192}$$

La formule (4.192) a été obtenue dans [17].

Ainsi, indépendamment de la forme de la distribution a priori du paramètre vectoriel du signal, sa distribution a posteriori, pour un échantillon de taille importante, est approximativement normale, de vecteur des moyennes égal au vecteur des estimations du maximum de vraisemblance et de matrice de corrélation $(1/n)\,\mathbf{I}^{-1}(\hat{\vartheta}_{mv})$.

**4.5.5. Propriétés asymptotiques des estimations de Bayes du paramètre vectoriel d'un signal pour des classes larges des bruits et des fonctions de pertes.** Comme on sait (voir tome II, 2.4.5), le risque a posteriori lors de l'estimation d'un paramètre vectoriel est

$$J(\hat{\vartheta} \mid x) = \int_{\Theta} \Pi(\hat{\vartheta} - \vartheta)\, W(\vartheta \mid x)\, d\vartheta. \tag{4.193}$$

L'estimation de Bayes $\hat{\vartheta}^*$ du paramètre $\vartheta$ est donnée par la condition du minimum du risque conditionnel :

$$J(\hat{\vartheta}^* \mid x) = \min_{\hat{\vartheta}} J(\hat{\vartheta} \mid x).$$

Supposons que la fonction de pertes $\Pi(\hat{\vartheta} - \vartheta)$ appartienne à la classe des fonctions symétriques non négatives telles que le minimum de l'intégrale

$$\int_{E_m} \Pi(u) \exp\left\{-\frac{1}{2}(u - \lambda)' \mathbf{H}(u - \lambda)\right\} du$$

est obtenu pour $\lambda = 0$ ($E_m$ est ici l'espace euclidien à $m$ dimensions, $\mathbf{H}$ une matrice définie positive de dimension $m \times m$). Telle est la fonction quadratique de pertes par exemple (voir tome II, 2.4.5).

Désignons par $g(\vartheta, \vartheta_0)$ le membre de droite de l'égalité (4.192) et, tout en conservant la notation (4.191), introduisons encore la grandeur

$$(\hat{\vartheta} - \hat{\vartheta}_{mv}) \sqrt{n} = \delta_n. \tag{4.194}$$

Pour la classe considérée des fonctions de pertes on a alors

$$\psi_n(\delta_n, \vartheta_0) = \int_{E_m} \Pi(\delta_n - \bar{\vartheta})\, g(\bar{\vartheta}, \vartheta_0)\, d\bar{\vartheta} =$$

$$= \int_{E_m} \Pi(u)\, g(u - \delta_n\,;\, \vartheta_0)\, du \geqslant \psi_n(0\,;\, \vartheta_0). \tag{4.195}$$

Montrons que pour $n \to \infty$ on a

$$\Phi_n(\delta_n, x) = \int_{E_m} \Pi(\delta_n - \bar{\vartheta})\, g_n(\bar{\vartheta} \mid x)\, d\bar{\vartheta} \geqslant$$

$$\geqslant \int_{E_m} \Pi(\bar{\vartheta})\, g_n(\bar{\vartheta} \mid x)\, d\bar{\vartheta} = \Phi_n(0\,;\, x), \qquad (4.196)$$

où

$$g_n(\bar{\vartheta} \mid x) = W[(\vartheta - \hat{\vartheta}_{mv})\sqrt{n} \mid x]. \qquad (4.197)$$

A partir de (4.195) et (4.196) on a

$$2 \mid \Phi_n(\delta_n, x) - \psi_n(\delta_n, \vartheta_0) \mid \leqslant 2 \int_{E_m} \Pi(\delta_n - \bar{\vartheta}) \mid g_n(\bar{\vartheta} \mid x) -$$

$$- g(\bar{\vartheta}, \vartheta_0) \mid d\bar{\vartheta} = \varepsilon_n(x, \vartheta_0), \qquad (4.198)$$

$$\Phi_n(\delta_n, x) \geqslant \Phi_n(0, x) - \varepsilon_n(x, \vartheta_0). \qquad (4.199)$$

Comme pour $n \to \infty$ en vertu de (4.192) on a $\varepsilon_n(x, \vartheta_0) \to 0$, ceci démontre l'inégalité (4.196). Cette inégalité signifie que pour $n \to \infty$ le risque a posteriori pour une estimation quelconque ne sera pas inférieur à celui obtenu pour l'estimation du maximum de vraisemblance *).

Ainsi, en utilisant l'estimation du maximum de vraisemblance qu'on peut trouver sans connaître à l'avance la fonction de pertes, ni la distribution des paramètres, on obtient pour un échantillon de taille suffisamment grande (temps d'observation) à peu près la même valeur du risque moyen que lorsque l'on utilise l'estimation de Bayes, pour une classe assez large de fonctions de pertes et des densités de probabilité a priori des paramètres. Cette propriété asymptotique des estimations de Bayes permet, dans de nombreux cas, de surmonter l'indétermination a priori, à condition d'utiliser pour le calcul des estimations des échantillons de tailles importantes. La densité a priori et la fonction de pertes inconnues peuvent être données de sorte à simplifier les calculs (réalisation de l'algorithme d'estimation). Au contraire, si l'estimation du maximum de vraisemblance est relativement facile à trouver (sous forme explicite ou par des relations de récurrence) on peut en rester là et renoncer à l'estimation de Bayes.

## 4.6. PROBLÈMES

**4.1.** Soit X un vecteur aléatoire distribué suivant une loi normale à $N$ dimensions, de vecteur des moyennes nul et de matrice de corrélation M. Montrer que la variable aléatoire $X'M^{-1}X$ est répartie suivant la loi du $\chi^2$ à $N$ degrés de liberté.

*) Il est facile de voir que la fonction $\Phi_n$ est un risque a posteriori exprimé à l'aide de variables normées.

**4.2.** Soient deux échantillons vectoriels d'apprentissage classifiés $\mathbf{x}_1^{(1)}, \ldots, \mathbf{x}_{n_1}^{(1)}$ et $\mathbf{x}_1^{(2)}, \ldots, \mathbf{x}_{n_2}^{(2)}$ de deux distributions normales à $N$ dimensions de mêmes matrices de corrélation $\mathbf{M}$ et de vecteurs des moyennes inconnus. On teste l'hypothèse sur l'égalité des valeurs moyennes de ces distributions. Etudier l'algorithme

$$\frac{n_1 n_2}{n_1+n_2}(\hat{\mathbf{a}}_1-\hat{\mathbf{a}}_2)'\,\mathbf{M}'\,(\hat{\mathbf{a}}_1-\hat{\mathbf{a}}_2) \geqslant \chi_\alpha^2(N) \tag{1}$$

en vertu duquel on rejette cette hypothèse. Dans la formule (1)

$$\hat{\mathbf{a}}_1=\frac{1}{n_1}\sum_{i=1}^{n_1}\mathbf{x}_i^{(1)}, \quad \hat{\mathbf{a}}_2=\frac{1}{n_2}\sum_{i=1}^{n_2}\mathbf{x}_i^{(2)} \tag{2}$$

et $\alpha$ est la probabilité donnée de rejeter par erreur l'hypothèse lorsqu'elle est vraie.

**4.3.** Soient deux classes caractérisées par des distributions normales unidimensionnelles de mêmes moyennes $a$ inconnues et de variances respectives $\sigma_1^2$ et $\sigma_2^2 = \lambda\sigma_1^2$ données. Etudier la règle de classification d'après l'observation $X$, basée sur le rapport de vraisemblance, où la moyenne inconnue est remplacée par l'estimation du maximum de vraisemblance obtenue à partir de l'échantillon commun, formé par deux échantillons d'apprentissage classifiés $x_1^{(1)}, \ldots, x_{n_1}^{(1)}$ et $x_1^{(2)}, \ldots, x_{n_2}^{(2)}$. Vérifier que cette estimation du maximum de vraisemblance est

$$\hat{a}=\frac{n_1\hat{a}_1+(n_2/\lambda)\,\hat{a}_2}{n_1+(n_2/\lambda)}, \tag{3}$$

où

$$\hat{a}_1=\frac{1}{n_1}\sum_{k=1}^{n_1}x_k^{(1)}, \quad \hat{a}_2=\frac{1}{n_2}\sum_{k=1}^{n_2}x_k^{(2)}, \tag{4}$$

et que

$$M_2\{\hat{a}\}=\frac{\sigma_1^2}{n_1+n_2/\lambda}. \tag{5}$$

Montrer de plus que l'estimation sans biais obtenue à partir de (3) en posant $\lambda = 1$, bien qu'elle ne soit pas une estimation du maximum de vraisemblance, est asymptotiquement normale et asymptotiquement efficace.

**4.4.** Généraliser l'algorithme de classification (4.7) au cas de $m$ classes $S_1, \ldots, S_m$ caractérisées par des distributions normales à $N$ dimensions, ayant les mêmes matrices de corrélation $\mathbf{M}_i = \mathbf{M}$ et les vecteurs des moyennes $\mathbf{a}_i$ différents, $i = 1, \ldots, m$. Dans le cas où $\mathbf{a}_i$ et $\mathbf{M}$ sont inconnus, et que l'on dispose des échantillons d'apprentissage classifiée $\mathbf{x}_1^{(i)}, \ldots, \mathbf{x}_{n_i}^{(i)}$, $i = 1, \ldots, m$, étudier

l'algorithme de classification suivant : l'observation X se rapporte à la classe $j$ si

$$\hat{V}_{jk} = \left(\mathbf{X} - \frac{\hat{\mathbf{a}}_l + \hat{\mathbf{a}}_k}{2}\right)' \hat{\mathbf{M}}^{-1} (\hat{\mathbf{a}}_l - \hat{\mathbf{a}}_k) \geqslant c_{jk}, \quad k = 1, \ldots, m; \; k \neq j, \tag{6}$$

où

$$\hat{\mathbf{a}}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \mathbf{x}_i^{(k)}, \quad k = 1, \ldots, m, \tag{7}$$

$$\hat{\mathbf{M}}^{-1} = \left(\sum_{i=1}^{m} n_i - m\right)^{-1} \sum_{i=1}^{m} \sum_{k=1}^{n_i} (\mathbf{x}_k^{(i)} - \hat{\mathbf{a}}_i)(\mathbf{x}_k^{(i)} - \hat{\mathbf{a}}_i)'. \tag{8}$$

Justifier le choix des seuils $c_{jk}$.

4.5. [29]. Analyser l'exemple de 4.3.6 en se donnant la distribution

$$w_1(\vartheta) = \alpha e^{-\alpha\vartheta}, \quad \alpha > 0, \quad \vartheta \geqslant 0, \tag{9}$$

et en substituant au lieu de $\alpha$ l'estimation $\hat{\alpha}$ du maximum de vraisemblance obtenue à partir de l'échantillon d'apprentissage $x_1, \ldots, x_n$.

On conseille de procéder à l'analyse en quatre étapes :

1) montrer que

$$\Delta(x, \alpha) = \alpha\,[\vartheta_0(1+\alpha) - (x+1)]/(1+\alpha)^{x+2}, \tag{10}$$

2) trouver l'expression de la fonction de vraisemblance

$$P(x_1, \ldots, x_n \mid \alpha) = \frac{\alpha}{(\alpha+n)^{1+\sum_{k=1}^{n} x_k}} \frac{\Gamma\left(\sum_{k=1}^{n} x_k\right)}{\prod_{k=1}^{n} x_k\,!}, \tag{11}$$

3) trouver l'estimation du maximum de vraisemblance du paramètre $\alpha$ :

$$\hat{\alpha} = \left(\frac{1}{n} \sum_{i=1}^{n} x_i\right)^{-1}, \tag{12}$$

4) écrire la statistique utilisée pour prendre une décision

$$\Delta_1(x, \hat{\alpha}) = \vartheta_0 \frac{1 + \frac{1}{n} \sum_{i=1}^{n} x_i}{\frac{1}{n} \sum_{i=1}^{n} x_i} - (x+1) \tag{13}$$

et analyser l'algorithme correspondant.

CHAPITRE 5

# PRINCIPES DE SIMILITUDE ET D'INVARIANCE DANS LA SYNTHÈSE STATISTIQUE

## 5.1. PRINCIPES DE SIMILITUDE ET D'INVARIANCE

**5.1.1. Rôle des principes de similitude et d'invariance dans la théorie de la détection des signaux.** Jetons un coup d'œil rétrospectif sur le contenu des chapitres précédents. Dans le chapitre 2 nous avons étudié les méthodes *non paramétriques* de synthèse des algorithmes de détection des signaux noyés dans des bruits. La limitation essentielle imposée aux éléments d'échantillons était qu'ils soient indépendants. Lorsque les observations sont corrélées les propriétés non paramétriques des algorithmes disparaissent. Les algorithmes asymptotiquement optimaux de détection examinés au chapitre 3, bien qu'applicables dans le cas de bruits corrélés, ne sont suffisamment efficaces que lorsque l'intervalle de temps d'observation est assez grand, de plus on voit apparaître ici des difficultés d'étude analytique de la vitesse de leur convergence vers l'algorithme optimal. Tous les algorithmes adaptatifs évoqués au chapitre 4 exigent un apprentissage, donc un temps supplémentaire.

Il existe des principes de statistique mathématique permettant de réaliser la synthèse des algorithmes de détection des signaux noyés dans des bruits corrélés sans apprentissage, algorithmes qui soient optimaux et stables non pas asymptotiquement, mais pour un temps d'observation fini (échantillon de taille finie). Ce sont les principes d'invariance et de similitude mentionnés au chapitre 1 (voir 1.2.3). Cependant l'utilisation de ces principes est également soumise à certaines restrictions. Tout d'abord ils sont applicables dans les conditions d'indétermination a priori paramétrique. Puis, le principe de similitude n'est efficace qu'avec la classe des distributions exponentielles des statistiques suffisantes. Cependant, même sous les restrictions mentionnées, le recours à des principes de similitude et d'invariance peut être utile dans nombre de problèmes importants de la radiodétection et des télécommunications. Nous allons envisager l'application de ces principes à la synthèse des algorithmes de détection des signaux noyés dans des bruits normaux à paramètres (perturbateurs) inconnus. Dans certains cas les paramètres perturbateurs seront les puissances inconnues des bruits, et dans d'autres les

paramètres des spectres énergétiques des bruits. Préalablement nous allons donner (sans démonstration) certains résultats indispensables de la théorie de la similitude et de l'invariance.

**5.1.2. Famille complète des densités de probabilité.** Considérons une famille uniparamétrique de densités de probabilité $w_1(x \mid \vartheta)$, $\vartheta \in \Theta$. Cette famille est dite complète si pour tous les $\vartheta \in \Theta$ l'équation intégrale par rapport à $f(x)$

$$m_1\{f(x)\} = \int_{-\infty}^{\infty} f(x)\, w_1(x \mid \vartheta)\, dx = 0 \tag{5.1}$$

a pour solution unique $f(x) \equiv 0$. Autrement dit, pour une famille complète $w_1(x \mid \vartheta)$, $\vartheta \in \Theta$, aucune fonction $f(x)$ non identiquement nulle ne peut être orthogonale à toutes les fonctions de cette famille.

A titre d'exemple d'une famille complète on peut citer la famille des densités de probabilité exponentielles (pour $\vartheta$ quelconque réel)

$$w_1(x \mid \vartheta) = \exp[\vartheta x + C(x) + D(\vartheta)], \quad -\infty \leqslant x \leqslant \infty, \tag{5.2}$$

où $D(\vartheta)$ est donné par les conditions de normalisation. En substituant (5.2) dans (5.1), on trouve que pour tous les $\vartheta$ appartenant à un certain intervalle on a

$$\int_{-\infty}^{\infty} \{f(x) \exp[C(x)]\}\, e^{\vartheta x}\, dx = 0,$$

p'où l'on tire $f(x) \exp[C(x)] \equiv 0$ et $f(x) \equiv 0$ car la transformation de Fourier est unique, donc à l'image nulle ne peut correspondre qu'une fonction de départ nulle.

La famille de densités de probabilité normales $w_1(x \mid \vartheta) = (2\pi\sigma^2)^{-1/2} e^{-(x-\vartheta)^2/(2\sigma^2)}$ pour une variance donnée $\sigma^2 = 1$ est un cas particulier de la famille (5.2) pour $C(x) = \ln\left[\frac{1}{\sqrt{2\pi}} e^{-x^2/2}\right]$, $D(\vartheta) = -\vartheta^2/2$.

Notons cependant que pour une valeur moyenne fixée (par exemple $\vartheta = 0$) la famille de densités normales $w_1(x \mid \sigma^2)$ n'est pas complète car

$$\int_{-\infty}^{\infty} f(x)\, e^{-x^2/(2\sigma^2)}\, dx = 0,$$

pour toute fonction $f(x)$ impaire non identiquement nulle.

Soit $\mathbf{x} = (x_1, \ldots, x_n)$ un échantillon de la distribution $w_1(x \mid \vartheta)$ et soit $t(\mathbf{x})$, $t \in T$, une statistique suffisante pour le paramètre $\vartheta$ de cette distribution. Désignons par $W_1(t \mid \vartheta)$ la densité de probabilité de la statistique suffisante. Si la famille $W_1(t \mid \vartheta)$ est complè-

te, la statistique suffisante $t(\mathbf{x})$ est dite complète, c'est-à-dire que la statistique est complète si telle est la famille des distributions à laquelle elle appartient. Supposons, par exemple, que l'échantillon indépendant $\mathbf{x}$ appartienne à une distribution normale de moyenne nulle. Il est facile de voir que la statistique suffisante pour l'estimation de la variance $\sigma^2$ est dans ce cas la grandeur $t(\mathbf{x}) = \sum_{i=1}^{n} x_i^2$ répartie suivant la loi du $\chi^2$ à $n$ degrés de liberté (voir tome II, page 49)

$$W_1(t \mid \sigma^2) = \left[2\sigma^n \Gamma\left(\frac{n}{2}\right)\right]^{-1} \left(\frac{t}{2}\right)^{(n/2)-1} e^{-t/(2\sigma^2)}, \quad t \geqslant 0. \quad (5.3)$$

La famille de fonctions (5.3) est complète car l'équation intégrale

$$\int_0^\infty f(t)\, t^{(n/2)-1} e^{-t/(2\sigma^2)}\, dt = 0$$

a une solution unique $f(t) = 0$ pour tous les $\sigma^2 > 0$. Par conséquent la statistique suffisante $t(\mathbf{x}) = \sum_{i=1}^{n} x_i^2$ est dans le cas envisagé complète.

Considérons une fonction $\varphi(\vartheta)$ du paramètre de la famille initiale $w_1(x \mid \vartheta)$ et supposons que pour estimer la fonction $\varphi(\vartheta)$ soit utilisée une fonction de la statistique suffisante $\Phi[t(\mathbf{x})]$. En mettant à profit le fait que cette estimation doit être sans biais, on obtient l'équation

$$\int_T \Phi(t)\, W_1(t \mid \vartheta)\, dt = \varphi(\vartheta). \quad (5.4)$$

Si $\Phi_1(t)$ et $\Phi_2(t)$ sont deux solutions de cette équation, la différence $f(t) = \Phi_1(t) - \Phi_2(t)$ doit satisfaire à l'équation

$$\int_T f(t)\, W_1(t \mid \vartheta)\, dt = 0.$$

Pour la statistique $\Phi(t)$, dont on sait qu'elle est complète, on a $f(t) = \Phi_1(t) - \Phi_2(t) = 0$, c'est-à-dire que si la statistique est complète, l'estimation non biaisée est obligatoirement unique.

Les résultats obtenus peuvent être généralisés au cas d'une famille multiparamétrique de distributions multidimensionnelles $w_N(\mathbf{x} \mid \boldsymbol{\vartheta})$ lorsque la variable $\mathbf{t}(\mathbf{x})$ et le paramètre $\boldsymbol{\vartheta}$ sont des vecteurs. Par exemple, est complète la famille exponentielle de distributions simultanées des statistiques suffisantes

$$W_m(\mathbf{t} \mid \boldsymbol{\vartheta}) = C(\boldsymbol{\vartheta}) \exp[\boldsymbol{\vartheta}'\mathbf{t}(\mathbf{x})] = C(\boldsymbol{\vartheta}) \exp\left[\sum_{i=1}^{m} \vartheta_i t_i(\mathbf{x})\right]. \quad (5.5)$$

**5.1.3. Algorithmes semblables de Neyman.** Revenons au problème statistique de la détection d'un signal de la forme $\lambda s(t, \vartheta)$, $\vartheta \in \Theta_s$, noyé dans un bruit dont la distribution dépend du vecteur des paramètres perturbateurs $\theta \in \Theta_b$. On vérifie l'hypothèse composée $H$: $\lambda = 0$, $\theta \in \Theta_b$, contre l'alternative composée $K$: $\lambda > 0$, $\vartheta \in \Theta_s$, $\theta \in \Theta_b$. Pour l'hypothèse $H$ la distribution des observations dépend des paramètres perturbateurs et ne dépend pas des paramètres informationnels (paramètres du signal). Comme nous l'avons noté dans 1.2.3, on peut essayer de trouver la classe de règles (algorithmes) semblables de vérification des hypothèses (détection d'un signal) qui présentent cette particularité que la probabilité de fausse alarme ne dépend pas des paramètres perturbateurs et conserve une valeur constante $\alpha$ pour toutes les distributions appartenant à la famille paramétrique des distributions du bruit $w_b(\mathbf{x} \mid \theta)$, $\theta \in \Theta_b$.

Supposons maintenant que le vecteur des statistiques suffisantes $\mathbf{t}(\mathbf{x})$ pour les paramètres gênants $\theta$ existe. Alors tout algorithme de détection utilisant la fonction de décision $\Phi(\mathbf{x})$ (voir (1.7)) pour laquelle, selon l'hypothèse $H$, la moyenne conditionnelle pour une valeur donnée $\mathbf{t}$ de la statistique suffisante est égale à une valeur donnée $\alpha$ de la probabilité de fausse alarme, sera semblable. En effet, à partir de la condition

$$m_1 \{\Phi(\mathbf{x}) \mid H, \mathbf{t}\} = \alpha, \tag{5.6}$$

en vertu de la propriété bien connue de la moyenne conditionnelle (voir tome I, page 78), on a

$$m_1 \{\Phi(\mathbf{x}) \mid H\} = m_1 \{m_1 \{\Phi(\mathbf{x}) \mid H, \mathbf{t}\}\} = \alpha. \tag{5.6'}$$

Les algorithmes satisfaisant à la condition (5.6) sont appelés algorithmes de Neyman pour la statistique $\mathbf{t}(\mathbf{x})$. La condition nécessaire et suffisante pour que les algorithmes semblables aient une structure de Neyman par rapport à la statistique suffisante $\mathbf{t}(\mathbf{x})$ est que la famille de distributions $W(\mathbf{t} \mid \theta)$ soit complète (voir [11]).

**5.1.4. Algorithmes optimaux de Neyman.** L'importance des algorithmes de Neyman réside dans le fait que la recherche, dans la classe de ces algorithmes, lorsque les paramètres du bruit sont inconnus, de l'algorithme optimal minimisant la probabilité de perte du signal pour une probabilité donnée de fausse alarme devient plus facile. Ceci est dû au fait que le problème formulé ci-dessus de vérification d'une hypothèse composée contre une alternative composée se réduit à la vérification d'une hypothèse simple contre une alternative composée (et parfois simple). En effet, les algorithmes de la structure de Neyman possèdent cette propriété que la probabilité de fausse alarme conserve une valeur constante $\alpha$ sur chacune des surfaces $\mathbf{t}(\mathbf{x}) = \mathbf{t}$. Aussi, peut-on considérer comme constante la statistique $\mathbf{t}(\mathbf{x})$, suffisante pour les paramètres perturbateurs, par con-

séquent la fonction de vraisemblance n'en dépend plus et l'hypothèse composée $H$ se ramène à une hypothèse simple $H_0$. Si de plus la statistique $\mathbf{t}(\mathbf{x})$ est une statistique suffisante des paramètres perturbateurs pour l'alternative également, l'alternative composée pour $\mathbf{t}$ constant devient une alternative simple (dans ce cas ni la probabilité de fausse alarme, ni la probabilité de perte du signal ne dépendent des paramètres perturbateurs du bruit).

Lors de la vérification d'une hypothèse simple contre une alternative simple la règle optimale impose, comme on sait, de comparer à un seuil le rapport de vraisemblance. Lorsque l'alternative est composée, on recherche la règle uniformément la plus puissante, ou la règle non biaisée uniformément la plus puissante (voir tome II, 1.4.4).

Soient $U(\mathbf{x})$ une statistique suffisante pour le paramètre informationnel $\vartheta$, et $\mathbf{t}(\mathbf{x})$ une statistique vectorielle suffisante pour les paramètres perturbateurs $\theta = (\theta_1, \ldots, \theta_m)$, et supposons que le rapport de vraisemblance dans le problème formulé ci-dessus de la détection d'un signal s'écrive comme suit :

$$l(u, \mathbf{t}) = C(\vartheta, \theta) \exp\left[\vartheta u + \sum_{i=1}^{m} \theta_i t_i\right]. \tag{5.7}$$

Le rapport de vraisemblance conditionnel pour $\mathbf{t}$ fixé ne dépend que de la variable $u$. Il est montré dans [11] que dans ce cas la règle non biaisée uniformément la plus puissante de vérification de l'hypothèse $H: \vartheta = \vartheta_0$ contre l'alternative $K: \vartheta > \vartheta_0$ revient à la comparaison à un seuil de la statistique suffisante $U(\mathbf{x})$, le seuil $C(\mathbf{t})$ étant donné par l'équation (voir (5.6))

$$m_1\{\Phi(U, \mathbf{t}) \mid \mathbf{t}, H\} = \alpha \tag{5.8}$$

pour tous les $\mathbf{t}$.

Cependant, il est parfois difficile d'utiliser l'équation (5.8) pour trouver le seuil. La solution se simplifie notablement s'il existe une statistique $V = h(U, \mathbf{t})$ ne dépendant pas de $\mathbf{t}$ pour l'hypothèse $H$ et qui, pour chaque $\mathbf{t}$ fixé, est monotone par rapport à $U$. Dans ce cas la fonction de décision uniformément la plus puissante sera

$$\Phi(V) = \begin{cases} 1, & V \geqslant C, \\ 0, & V < C, \end{cases} \tag{5.9}$$

le seuil $C$ étant déterminé d'après la probabilité donnée $\alpha$ d'erreurs de première espèce par la relation

$$m_1\{\Phi(V) \mid H\} = \alpha. \tag{5.9'}$$

Il se peut qu'aucune fonction randomisée du type (5.9) ne vérifie la condition (5.8), pourtant cette condition est satisfaite par une

fonction randomisée

$$\Phi(V)=\begin{cases}1, & V>C,\\ \gamma, & V=C,\\ 0, & V<C,\end{cases} \tag{5.9''}$$

où les deux constantes, $C$ et $\gamma$, se trouvent à partir de l'équation (5.8).

**5.1.5. Règles invariantes de choix d'une décision lors de la vérification des hypothèses.** Dans certains cas, pour trouver une règle optimale de choix d'une décision lors de la vérification d'une hypothèse composée contre une alternative composée, on peut utiliser le principe de l'invariance par rapport à un certain groupe de transformations (voir 1.2.3).

Soit $G$ un groupe de transformations qui sont des applications biunivoques de l'espace échantillonné $\mathbf{X}$ sur lui-même. La famille paramétrique de distributions $W(\mathbf{x} \mid \boldsymbol{\vartheta})$, $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$, est dite invariante par rapport au groupe $G$, si à chaque application $g \in G$ correspond une application biunivoque $\bar{g} \in \bar{G}$ de l'ensemble $\boldsymbol{\Theta}$ sur lui-même. Ainsi, si $W(\mathbf{x} \mid \boldsymbol{\vartheta})$ est la distribution de la variable aléatoire $\mathbf{x}$, alors $W[\mathbf{y} \mid \bar{g}(\boldsymbol{\vartheta})]$ est la distribution de la variable aléatoire $\mathbf{y} = g(\mathbf{x})$. On dit que le problème de la vérification de l'hypothèse $H: \boldsymbol{\vartheta} \in \boldsymbol{\Theta}_H$ contre l'alternative $K: \boldsymbol{\vartheta} \in \boldsymbol{\Theta}_K$ reste invariant par rapport à la transformation $g$, si la transformation $\bar{g}$ qui lui correspond conserve les espaces $\boldsymbol{\Theta}_H$ et $\boldsymbol{\Theta}_K$, c'est-à-dire

$$\bar{g}\boldsymbol{\Theta}_H=\boldsymbol{\Theta}_H, \quad \bar{g}\boldsymbol{\Theta}_K=\boldsymbol{\Theta}_K. \tag{5.10}$$

La règle de choix d'une décision, lors de la vérification de l'hypothèse $H$ contre l'alternative $K$, est dite invariante par rapport au groupe $G$ si la fonction de décision $\Phi(\mathbf{x})$ pour tous les $\mathbf{x} \in \mathbf{X}$ et $g \in G$ satisfait à la relation

$$\Phi(g\mathbf{x}) = \Phi(\mathbf{x}). \tag{5.11}$$

Une fonction $\mathrm{M}(\mathbf{x})$, $\mathbf{x} \in \mathbf{X}$, est appelée *invariant maximal* par rapport au groupe de transformations $G$ si elle est invariante par rapport à ce groupe, c'est-à-dire si $\mathrm{M}[g(\mathbf{x})] = \mathrm{M}(\mathbf{x})$ pour tous les $\mathbf{x} \in \mathbf{X}$ et $g \in G$, et si $\mathrm{M}(\mathbf{x}_1) = \mathrm{M}(\mathbf{x}_2)$ entraîne $\mathbf{x}_2 = g(\mathbf{x}_1)$ pour un certain $g \in G$. L'importance de la notion introduite devient évidente à partir des considérations suivantes (voir [11]): la condition nécessaire et suffisante d'invariance d'une règle de choix d'une décision par rapport au groupe de transformations $G$ consiste en ce que la fonction de décision $\Phi(\mathbf{x})$ correspondant à cette règle ne dépende de $\mathbf{x}$ que par l'intermédiaire de l'invariant maximal $\mathrm{M}(\mathbf{x})$ par rapport à la transformation mentionnée, c'est-à-dire que

$$\Phi(\mathbf{x}) = \Psi[\mathrm{M}(\mathbf{x})]. \tag{5.12}$$

Ainsi, la classe des fonctions de décision invariantes peut être obtenue comme un ensemble de fonctions de l'invariant maximal.

De plus, il y a lieu de noter que l'invariance permet non seulement de réduire la dimension des données traitées, ceci en passant de l'espace des échantillons à l'invariant maximal, mais également de diminuer la dimension de l'espace paramétrique. L'affirmation correspondante se formule comme suit: si M(x) est invariante par rapport au groupe de transformations $G$ et que $v(\vartheta)$ soit l'invariant maximal par rapport au groupe $\overline{G}$ (voir (5.10)), la distribution M(x) ne dépend que de $v(\vartheta)$.

Pour une telle réduction des données le problème initial se réduit à la vérification d'une hypothèse simple, contre une alternative composée (et dans certains cas également simple). Si de plus la famille de densités de probabilité $W$ [M (x) | v ($\vartheta$)] a un rapport de vraisemblance monotone, la règle invariante uniformément la plus puissante de choix d'une décision est basée sur la comparaison à un seuil de la statistique M (x).

## 5.2. DÉTECTION D'UN SIGNAL NOYÉ DANS UN BRUIT NORMAL ADDITIF DE PUISSANCE INCONNUE

**5.2.1. Statistiques suffisantes dans le problème de la détection d'un signal déterministe.** Nous allons envisager le problème de la détection d'un signal déterministe $\mu s(t)$, $\mu > 0$, noyé dans un bruit normal additif $\xi(t)$ de valeur moyenne nulle et de fonction de corrélation $\sigma^2 R(t, y)$, la fonction $R(t, y)$ étant donnée et la grandeur $\sigma^2$ inconnue. Pour un bruit stationnaire la grandeur $\sigma^2$ est la variance ou la puissance moyenne.

En termes de la théorie de la vérification des hypothèses, le problème se formule comme suit: vérifier l'hypothèse composée $H$: $\mu = 0$, $\sigma^2 > 0$, contre l'alternative composée $K$: $\mu > 0$, $\sigma^2 > 0$. Ici $\mu$ est un paramètre informatif et $\sigma^2$ un paramètre perturbateur. A partir de la réalisation $x(t)$ observée sur l'intervalle $(-T, T)$, il y a lieu de choisir l'une des deux décisions:

$$\gamma_H : x(t) \in \xi(t) \text{ (l'hypothèse } H \text{ est vraie)},$$

$$\gamma_K : x(t) \in \mu s(t) + \xi(t),\ \mu > 0 \text{ (l'alternative } K \text{ est vraie)}.$$

Notons que dans le problème formulé on est dans l'incertitude tant par rapport à l'amplitude $\mu$ du signal (paramètre informatif), que par rapport à la variance du bruit $\sigma^2$ (paramètre perturbateur).

La règle optimale de choix d'une décision pour $\sigma^2$ donnée a été étudiée dans le tome II. Lorsque $\sigma^2$ est inconnue, il n'existe pas de règle uniformément la plus puissante, et on peut essayer de trouver une règle optimale dans la classe de règles soumises à des restrictions qui ont été énumérées dans 5.1.

Nous allons avant tout passer de la réalisation continue $x(t)$, observée sur l'intervalle $(-T, T)$, aux coordonnées $x_k$, $k = 1, \ldots, N$, indépendantes normalement réparties, obtenues par le développement orthogonal du processus aléatoire normal (voir tome II, 5.2)

$$x_k = \sqrt{\lambda_k} \int_{-T}^{T} x(t)\, \varphi_k(t)\, dt, \tag{5.13}$$

où $\lambda_k$ et $\varphi_k(t)$ sont les valeurs propres et les fonctions propres de l'équation intégrale

$$\varphi(t) = \lambda \int_{-T}^{T} R(t, y)\, \varphi(y)\, dy, \quad |t| \leqslant T. \tag{5.13'}$$

Il est évident que $m_1\{x_k\} = 0$ pour l'hypothèse $H$ tandis que pour l'alternative $K$ $m_1\{x_k\} = \mu s_k$, $\mu > 0$, où

$$s_k = \sqrt{\lambda_k} \int_{-T}^{T} s(t)\, \varphi_k(t)\, dt. \tag{5.14}$$

La variance de $x_k$ est égale à $\sigma^2$ tant pour l'hypothèse que pour l'alternative. Pour ne pas perdre une partie importante de l'information utile, lorsque l'on remplace la réalisation $x(t)$ par un échantillon fini $\mathbf{x} = (x_1, \ldots, x_N)$, il faut que $N$ soit suffisamment grand.

La distribution conjointe des variables aléatoires normales indépendantes $x_k$, $k = 1, \ldots, N$, pour $\mu > 0$ est donnée par la famille biparamétrique suivante:

$$W_N(\mathbf{x} \mid \mu, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{k=1}^{N} (x_k - \mu s_k)^2\right\}. \tag{5.15}$$

En introdisant les désignations

$$\vartheta_1 = \mu/\sigma^2, \quad \vartheta_2 = 1/(2\sigma^2), \tag{5.16}$$

$$U(\mathbf{x}) = \sum_{k=1}^{N} x_k s_k, \tag{5.17}$$

$$t(\mathbf{x}) = \sum_{k=1}^{N} x_k^2, \tag{5.18}$$

$$C(\vartheta_1, \vartheta_2) = (\vartheta_2/\pi)^{N/2} \exp\left(-\frac{\vartheta_1^2}{4\vartheta_2} \sum_{k=1}^{N} s_k^2\right), \tag{5.19}$$

on peut écrire (5.15) sous la forme

$$W_N(\mathbf{x} \mid \mu, \sigma^2) = C(\vartheta_1, \vartheta_2) \exp[\vartheta_1 U(\mathbf{x}) - \vartheta_2 t(\mathbf{x})]. \tag{5.20}$$

En vertu de (5.20) les statistiques $U(\mathbf{x})$ et $t(\mathbf{x})$ sont suffisantes pour les paramètres $\vartheta_1$, $\vartheta_2$. Le problème formulé ci-dessus de vérification des hypothèses équivaut à celui de la vérification de l'hypothèse composée $H: \vartheta_1 = 0, \vartheta_2 > 0$, contre l'alternative composée $K: \vartheta_1 > 0, \vartheta_2 > 0$. Maintenant c'est le paramètre $\vartheta_1$ qui est informatif, et le paramètre $\vartheta_2$ perturbateur.

**5.2.2. Règle sans biais uniformément la plus puissante de détection d'un signal déterministe.** A partir de (5.20) on obtient l'expression du logarithme du rapport de vraisemblance

$$\ln l(\mathbf{x} \mid \vartheta_1, \vartheta_2) = \ln \frac{W_N(\mathbf{x} \mid \mu, \sigma^2)}{W_N(\mathbf{x} \mid 0, \sigma^2)} = \vartheta_1 U(\mathbf{x}) - \frac{\vartheta_1^2}{4\vartheta_2} \sum_{k=1}^{N} s_k^2. \quad (5.21)$$

La formule (5.21) montre que le rapport de vraisemblance dans le problème envisagé de détection d'un signal est une fonction monotone de la statistique suffisante $U(\mathbf{x})$ du type (5.7). C'est pourquoi, comme nous l'avons mentionné dans 5.1.4, la règle sans biais uniformément la plus puissante de choix d'une décision impose la comparaison de $U(\mathbf{x})$ à un seuil dépendant de $t(\mathbf{x})$, à savoir on prend la décision $\gamma_K$ sur la présence du signal si

$$\sum_{k=1}^{N} x_k s_k \geqslant C[t(\mathbf{x})], \quad (5.22)$$

et la décision $\gamma_H$ que le signal est absent, si c'est l'inégalité inverse de (5.22) qui est vérifiée.

La recherche de la fonction de seuil $C[t(\mathbf{x})]$ est un problème plus compliqué. A cet effet, on peut se donner $t$ et exiger que le plan

$$\sum_{k=1}^{N} x_k s_k = C$$

divise la sphère $t(\mathbf{x}) = t$ en deux parties et que la surface de la partie où $\sum_{k=1}^{N} x_k s_k \geqslant C$ soit égale au produit de la surface de toute la sphère par la probabilité donnée de fausse alarme.

On peut résoudre ce problème en utilisant l'invariance de la règle par rapport à un certain groupe de transformations (ce qui sera montré dans 5.2.4).

**5.2.3. Détection d'un signal constant.** Commençons par le problème le plus simple de la détection d'un signal constant ($s(t) \equiv 1$) noyé dans un bruit additif normal de valeur moyenne nulle et de

variance inconnue. Dans ce cas la règle (5.22) s'écrit comme suit:

$$U(\mathrm{x}) = \frac{1}{N} \sum_{k=1}^{N} x_k \geqslant C\,[t(\mathrm{x})], \tag{5.23}$$

où la statistique $t(\mathrm{x})$ est donnée par (5.18).

La règle (5.23) de choix d'une décision peut s'écrire sous une forme plus simple en tenant compte des considérations exposées à la fin du p. 5.1.4. Introduisons la statistique

$$V(\mathrm{x}) = \frac{U(\mathrm{x})\sqrt{N/(N-1)}}{\sqrt{t(\mathrm{x}) - NU^2(\mathrm{x})}} = \frac{U(\mathrm{x})\sqrt{N/(N-1)}}{\sqrt{\sum_{k=1}^{N} [x_k - U(\mathrm{x})]^2}}. \tag{5.24}$$

Pour l'hypothèse $H$: $\mu = 0$ (il n'y a pas de signal) cette statistique ne dépend pas de $t(\mathrm{x}) = \sum_{i=1}^{N} x_i^2$ (voir [11], page 221) et pour chaque $t$ fixé est monotone par rapport à $U$. En vertu de (5.9), la règle sans biais uniformément la plus puissante se formule comme suit: le signal est présent si

$$V(\mathrm{x}) \geqslant C, \tag{5.25}$$

et l'on prend la décision qu'il n'y a pas de signal dans le cas contraire.

Pour l'hypothèse $H$ la statistique $V(\mathrm{x})$ est le rapport de deux variables aléatoires indépendantes: de la variable normale $(1/\sqrt{N}) \sum_{i=1}^{N} x_k/\sigma$ de moyenne nulle et de variance unité par la racine carrée de la variable aléatoire $[(N-1)\,\sigma^2]^{-1} \sum_{k=1}^{N} \left(x_k - \frac{1}{N}\sum_{i=1}^{N} x_i\right)^2$ répartie suivant la loi du $\chi^2$ à $N-1$ degrés de liberté, divisée par $N-1$. Donc la statistique $V(\mathrm{x})$ est répartie suivant la loi type de Student et par conséquent le seuil $C$ dans (5.25) pour une valeur $\alpha$ donnée de fausse alarme est égal à l'$\alpha$-quantile $t_\alpha$ de la distribution de Student.

Ainsi, l'inégalité (5.25) peut s'écrire comme suit (comparer avec (1.161) du tome II)

$$\frac{1}{\sqrt{N}} \sum_{k=1}^{N} x_k \geqslant \left\{\frac{1}{N-1} \sum_{k=1}^{N} \left(x_k - \frac{1}{N}\sum_{i=1}^{N} x_i\right)^2\right\}^{1/2} t_\alpha. \tag{5.26}$$

La probabilité de perte du signal s'écrit sous la forme (voir tome II, (1.162))

$$\beta = \int_{-\infty}^{t_\alpha} s_N(y, \delta_N)\, dy, \tag{5.27}$$

où $s_N(y, \delta_N)$ est la densité de probabilité de la distribution de Student non centrée de paramètre d'écart égal à

$$\delta_N = \frac{\mu}{\sigma}\sqrt{N}. \tag{5.27'}$$

**5.2.4. Utilisation du principe d'invariance.** Revenons au problème général de la détection d'un signal déterministe noyé dans un bruit additif normal de puissance moyenne inconnue. Pour tourner les difficultés liées au calcul de la fonction de seuil dans l'algorithme (5.22), essayons de voir s'il y a un groupe de transformations par rapport auquel, dans le cas envisagé, il existe une règle invariante de choix d'une décision.

Introduisons préalablement une nouvelle normalisation du signal de telle sorte que $\sum_{k=1}^{N} s_k^2 = 1$. Au lieu de l'amplitude $\mu$ nous allons considérer l'amplitude $\mu_N$ dépendant du nombre de coordonnées envisagées. Nous pouvons poser à cet effet

$$\mu_N^2 = \mu^2 \sum_{k=1}^{N} s_k^2. \tag{5.28}$$

Ainsi normée (voir (5.17)), la statistique

$$U_1(\mathrm{x}) = \frac{1}{\sigma} U(\mathrm{x}) = \frac{1}{\sigma} \sum_{k=1}^{N} x_k s_k \tag{5.29}$$

est répartie suivant une loi normale de paramètres $(\mu_N/\sigma, 1)$, et la statistique (voir (5.18))

$$U_2(\mathrm{x}) = \frac{1}{\sigma^2}[t(\mathrm{x}) - U^2(\mathrm{x})] = \frac{1}{\sigma^2}\left[\sum_{k=1}^{N} x_k^2 - \left(\sum_{k=1}^{N} x_k s_k\right)^2\right] =$$

$$= \frac{1}{\sigma^2}\sum_{k=1}^{N}(x_k - \mu_N s_k)^2 - \frac{1}{\sigma^2}\left(\sum_{k=1}^{N} x_k s_k - \mu_N\right)^2 \tag{5.29'}$$

suivant la loi du $\chi^2$ à $N-1$ degrés de liberté. Les statistiques $U_1(\mathrm{x})$ et $U_2(\mathrm{x})$ étant indépendantes, leur densité de probabilité simultanée est égale au produit des densités de probabilité de chacune des statistiques, soit

$$W_2(u_1, u_2) = \left(1/\sqrt{2\pi}\right)\exp\left[-(u_1 - \mu_N/\sigma)^2/2\right] \times$$
$$\times u_2^{(N-3)/2} e^{-u_2/2}\left[2^{(N-1)/2}\Gamma\left(\frac{N-1}{2}\right)\right]^{-1},$$
$$-\infty < u_1 < \infty, \quad 0 \leqslant u_2 < \infty. \tag{5.30}$$

En utilisant (5.29), (5.30) on obtient la loi simultanée des statistiques *liées* $U$ (x) et $t$ (x):

$$W_2(U, t \mid \vartheta_1, \vartheta_2) = |J| \left[\sqrt{2\pi}\, 2^{(N-1)/2} \Gamma\left(\frac{N-1}{2}\right)\right]^{-1} \times$$

$$\times (2\vartheta_2)^{(N-3)/2} e^{-\vartheta_1^2/(4\vartheta_2)} (t-U)^{(N-3)/2} \exp(\vartheta_1 U - \vartheta_2 t), \quad (5.31)$$

où

$$\vartheta_1 = \mu_N/\sigma^2, \ \vartheta_2 = 1/(2\sigma^2), \quad (5.31')$$

et le jacobien $J$ de la transformation est

$$J = \frac{\partial(u_1, u_2)}{\partial(U, t)} = \frac{\partial u_1}{\partial U} \frac{\partial u_2}{\partial t} = 1/\sigma^3. \quad (5.31'')$$

Introduisons les désignations:

$$C(\vartheta_1, \vartheta_2) = \left[\sqrt{2\pi}\, 2^{(N-1)/2} \Gamma\left(\frac{N-1}{2}\right)\right]^{-1} (2\vartheta_2)^{N/2} e^{-\vartheta_1^2/(4\vartheta_2)}, \quad (5.32)$$

$$h(U, t) = (t - U^2)^{(N-3)/2}; \quad (5.32')$$

nous pouvons maintenant écrire (5.31) comme suit:

$$W(u, t \mid \vartheta_1, \vartheta_2) = C(\vartheta_1, \vartheta_2)\, h(U, t) \exp(\vartheta_1 U - \vartheta_2 t),$$
$$-\infty < U < \infty, \ 0 \leqslant t < \infty. \quad (5.33)$$

Introduisons le groupe de transformations

$$g(U, t) = (aU, a^2 t), \ a > 0. \quad (5.34)$$

Toute transformation de ce groupe est une multiplication de la réalisation observée par une certaine constante $a$. Montrons que le problème envisagé de la détection est invariant par rapport au groupe de transformations (5.34). On tire de (5.33)

$$W[g(u, t)] = C[\bar{g}(\vartheta_1, \vartheta_2)]\, h(u, t) \exp[a\vartheta_1 U - a^2 \vartheta_2 t), \quad (5.35)$$

où

$$\bar{g}(\vartheta_1, \vartheta_2) = (a\vartheta_1, a^2\vartheta_2), \quad a > 0. \quad (5.36)$$

Comme dans notre problème $\Theta_H = (\vartheta_1 = 0, \vartheta_2 > 0)$, $\Theta_K = (\vartheta_1 > 0, \vartheta_2 > 0)$, il est évident que la transformation (5.36) satisfait à la relation (5.10), ce qui démontre l'invariance.

Montrons maintenant que la statistique

$$\mathrm{M}(\mathrm{x}) = U(\mathrm{x}) \{[t(\mathrm{x}) - U^2(\mathrm{x})]/(N-1)\}^{-1/2} \quad (5.37)$$

est l'invariant maximal par rapport au groupe de transformations (5.34). L'invariance de cette statistique est évidente car $\mathrm{M}(a\mathrm{x}) = \mathrm{M}(\mathrm{x}) = \mathrm{M}[U(\mathrm{x}), t(\mathrm{x})]$. D'autre part

$$\mathrm{M}[U(\mathrm{x}_1), t(\mathrm{x}_1)] = \mathrm{M}[U(\mathrm{x}_2), t(\mathrm{x}_2)]$$

entraîne

$$U(\mathbf{x}_1) = aU(\mathbf{x}_2),\ t(\mathbf{x}_1) = a^2 t(\mathbf{x}_2),\ a > 0$$

et

$$\mathbf{x}_1 = a\mathbf{x}_2,\ a > 0.$$

L'invariant maximal pour le groupe $\bar{g}(\vartheta_1, \vartheta_2)$ est la grandeur $\mu_N/\sigma$. L'invariant maximal (5.37) est réparti suivant la loi de Student non centrée $s_N(\mu, \delta_N)$ de paramètre d'écart $\delta_N = \mu_N/\sigma$. Mais comme le rapport de vraisemblance

$$l(\mathrm{M}) = s_N(\mathrm{M}, \delta_N)/s_N(\mathrm{M}, 0) \tag{5.38}$$

est une fonction monotone de M (voir [11], page 303), pour le problème envisagé de la détection d'un signal déterministe on a la règle

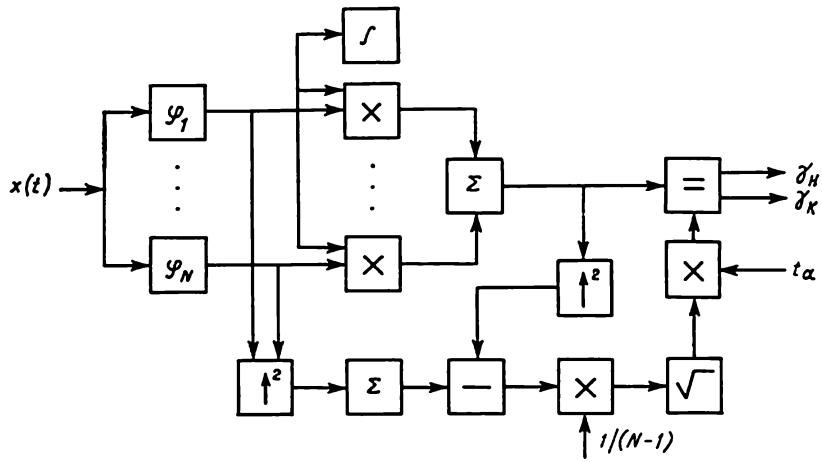



Fig. 5.1. Schéma fonctionnel du détecteur invariant d'un signal noyé dans un bruit additif normal de variance inconnue.

suivante uniformément la plus puissante de choix d'une décision : le signal est présent si

$$\mathrm{M}(\mathbf{x}) = U(\mathbf{x})\{[t(\mathbf{x}) - U^2(\mathbf{x})]/(N-1)\}^{-1/2} \geqslant t_\alpha, \tag{5.39}$$

où $t_\alpha$ est l'$\alpha$-quantile de la loi (centrale) de Student.

**5.2.5. Structure et caractéristique de fonctionnement du détecteur d'un signal déterministe.** Ecrivons l'inégalité (5.39) comme suit:

$$\sum_{k=1}^{N} x_k s_k \geqslant t_\alpha \left\{\left[\sum_{k=1}^{N} x_k^2 - \left(\sum_{k=1}^{N} x_k s_k\right)^2\right]/(N-1)\right\}^{1/2}. \tag{5.40}$$

Le schéma fonctionnel du détecteur corréspondant à l'algorithme (5.40) est donné sur la figure 5.1.

Le membre de gauche de l'inégalité (5.40) est la somme pondérée des coordonnées observées du processus. Pour une variance donnée d'un bruit normal additif, l'algorithme optimal de détection d'un signal déterministe prescrit la comparaison de cette somme à un seuil constant dépendant d'une probabilité donnée de fausse alarme (voir tome II, (5.6)). Lorsqu'on ne connaît pas la variance, la somme pondérée est comparée au seuil

$$C(\mathrm{x}) = t_\alpha \{[\sum_{k=1}^{N} x_k^2 - (\sum_{k=1}^{N} x_k s_k)^2]/(N-1)\}^{1/2} \qquad (5.40')$$

qui est le produit de l'$\alpha$-quantile de la distribution de Student par l'estimation sans biais de la variance, calculée d'après l'échantillon observé. Ainsi le détecteur donné sur la figure 5.1 peut être interprété comme un dispositif adaptatif utilisant l'algorithme optimal de détection d'un signal pour une variance connue du bruit, où

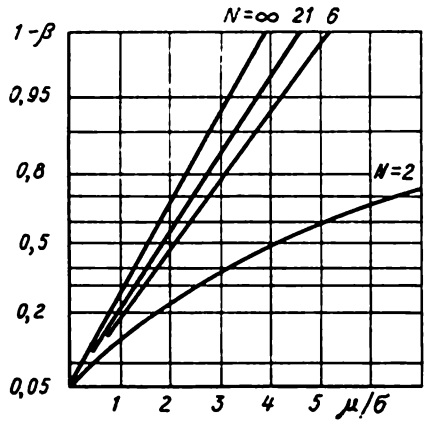



Fig. 5.2. Caractéristiques de fonctionnement du détecteur représenté sur la figure 5.1.

l'on a remplacé la variance par son estimation sans biais calculée d'après un échantillon non classifié. Lorsque $N \to \infty$, l'estimation $\hat{\sigma}^2 = [\sum_{k=1}^{N} x_k^2 - (\sum_{k=1}^{N} x_k s_k)^2]/(N-1)$ converge en probabilité vers $\sigma^2$ tant pour l'hypothèse que pour l'alternative, et à partir de (5.40) on obtient l'algorithme optimal de détection d'un signal noyé dans un bruit additif normal de variance connue.

La figure 5.2 reproduit les caractéristiques de fonctionnement, empruntées à [19], du détecteur représenté sur la figure 5.1. Ces caractéristiques donnent la probabilité de détection correcte en fonction du rapport $\mu/\sigma$ signal/bruit. Les courbes sont construites pour une

probabilité donnée de fausse alarme $\alpha = 0{,}05$ pour trois valeurs de $N = 2, 6, 21$. La courbe supérieure ($N = \infty$) correspond à l'algorithme optimal pour une variance donnée du bruit.

La caractéristique de fonctionnement du détecteur ne dépendant que du rapport $\mu/\sigma$, on pourrait formuler le problème traité comme un problème de détection d'un signal déterministe noyé dans un bruit normal additif pour un rapport signal/bruit inconnu.

**5.2.6. Détection d'un signal à bande étroite noyé dans un bruit normal additif de puissance inconnue.** Considérons maintenant le problème de la détection d'un signal à bande étroite

$$\mu s(t, \varphi_0) = \mu a(t) \cos[\omega_0 t - \psi_s(t) + \varphi_0], \quad \mu > 0, \quad |\varphi_0| \leqslant \pi, \tag{5.41}$$

modulé en amplitude et en phase selon les lois $a(t)$ et $\psi_s(t)$ respectivement, d'amplitude $\mu$ et de phase initiale $\varphi_0$ inconnues, noyé dans un bruit normal additif $\xi(t)$ de valeur moyenne nulle, de fonction de corrélation normée donnée $R(t, y)$ et de variance inconnue $\sigma^2$. Dans ce problème le paramètre informatif est l'amplitude du signal $\mu$, et les paramètres perturbateurs sont la phase du signal $\varphi_0$ et la variance du bruit $\sigma^2$.

En termes de la théorie de la vérification des hypothèses, le problème se formule comme suit : vérifier l'hypothèse composée $H: \mu = 0$, $|\varphi_0| \leqslant \pi$, $\sigma^2 > 0$ contre l'alternative composée $K: \mu > 0$, $|\varphi_0| \leqslant \pi$, $\sigma^2 > 0$. L'observation de la réalisation $x(t)$ sur l'intervalle $(-T, T)$ doit permettre de choisir l'une des décisions : $\gamma_H: x(t) = \xi(t)$ ou $\gamma_K: x(t) = \mu s(t) + \xi(t)$.

Représentons sous forme complexe la réalisation d'un processus aléatoire à bande étroite :

$$x(t) = \operatorname{Re} z(t) e^{i\omega_0 t}, \tag{5.42}$$

où $z(t)$ est l'enveloppe complexe, de plus

$$z(t) = A(t) + iC(t), \tag{5.42'}$$

où $A(t)$, $C(t)$ sont les composantes en quadrature du processus aléatoire à bande étroite. La fonction du signal $s(t, \varphi_0)$ s'écrit également à l'aide d'une enveloppe complexe, soit

$$s(t, \varphi_0) = \operatorname{Re} z_s(t) e^{i\omega_0 t} e^{i\varphi_0}. \tag{5.43}$$

Pour un bruit additif normal, la réalisation $x(t)$ sur l'intervalle d'observation $(-T, T)$ peut être décrite par un ensemble de coordonnées normalement réparties de l'enveloppe complexe (voir

tome II, 5.5.2)

$$z_k = \sqrt{\lambda_k} \int_{-T}^{T} z(t) \overline{\varphi_k(t)}\, dt, \tag{5.44}$$

où $\lambda_k$, $\varphi_k(t)$ sont les valeurs propres ($\lambda_k > 0$) et les fonctions propres de l'équation intégrale

$$\varphi(t) = \lambda \int_{-T}^{T} R_z(t, u)\, \varphi(u)\, du, \tag{5.44'}$$

avec

$$\sigma^2 R_z(t, u) = \frac{1}{2} m_1 \{z(t) \overline{z(u)}\}, \tag{5.44''}$$

où $z(t)$ est l'enveloppe complexe du bruit normal de valeur moyenne nulle.

Si $x_k$ et $y_k$ sont les parties réelle et imaginaire de la coordonnée $z_k$, l'ensemble $(\mathbf{x}, \mathbf{y}) = (x_1, y_1, \ldots, x_N, y_N)$ est un ensemble de variables aléatoires normales indépendantes; par ailleurs, pour l'hypothèse $H$ on a

$$m_1\{x_k\} = m_1\{y_k\} = 0, \qquad k = 1, 2, \ldots, \tag{5.45}$$

et pour l'alternative $K$

$$m_1\{x_k\} = \mu(a_k \cos\varphi_0 - b_k \sin\varphi_0), \tag{5.46}$$

$$m_1\{y_k\} = \mu(a_k \sin\varphi_0 + b_k \cos\varphi_0), \tag{5.46'}$$

$$a_k + ib_k = \sqrt{\lambda_k} \int_{-T}^{T} z_s(t) \overline{\varphi_k(t)}\, dt. \tag{5.46''}$$

La variance de chacune des variables aléatoires $x_k$, $y_k$, $k = 1, 2, \ldots$, est égale à $\sigma^2$.

La distribution simultanée des vecteurs aléatoires normaux $\mathbf{x}$, $\mathbf{y}$ est donnée par la famille triparamétrique de densités de probabilité suivante :

$$W_{2N}(\mathbf{x}, \mathbf{y} \mid \mu, \varphi_0, \sigma^2) =$$
$$= (2\pi\sigma^2)^{-N} \exp\left\{-\frac{1}{2\sigma^2} \sum_{k=1}^{N} [x_k - \mu(a_k \cos\varphi_0 - b_k \sin\varphi_0)]^2 + \right.$$
$$\left. + [y_k - \mu(a_k \sin\varphi_0 + b_k \cos\varphi_0)]^2\right\}. \tag{5.47}$$

En introduisant les désignations

$$\vartheta_1 = \mu \cos \varphi_0/\sigma^2, \quad \vartheta_2 = \mu \sin \varphi_0/\sigma^2, \quad \vartheta_3 = 1/(2\sigma^2), \tag{5.48}$$

$$U_1(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^{N} (x_k a_k + y_k b_k), \quad U_2(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^{N} (y_k a_k - x_k b_k), \tag{5.49}$$

$$t(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^{N} (x_k^2 + y_k^2), \tag{5.50}$$

$$C(\vartheta_1, \vartheta_2, \vartheta_3) = \left(\frac{\vartheta_3}{\pi}\right)^N \exp\left\{-\frac{\vartheta_1^2+\vartheta_2^2}{4\vartheta_3} \sum_{k=1}^{N} (a_k^2 + b_k^2)\right\}, \tag{5.51}$$

on peut écrire (5.47) comme suit

$$W_{2N}(\mathbf{x}, \mathbf{y} \mid \mu, \varphi_0, \sigma^2) = \\ = C(\vartheta_1, \vartheta_2, \vartheta_3) \exp\{\vartheta_1 U_1(\mathbf{x}, \mathbf{y}) + \vartheta_2 U_2(\mathbf{x}, \mathbf{y}) - \vartheta_3 t(\mathbf{x}, \mathbf{y})\}. \tag{5.52}$$

Notre problème de vérification des hypothèses est équivalent au problème de la vérification de l'hypothèse composée $H$: $|\boldsymbol{\vartheta}| = 0$, $\vartheta_3 > 0$ contre l'alternative composée $K$: $|\boldsymbol{\vartheta}| \neq 0$, $\vartheta_3 > 0$, où le vecteur $\boldsymbol{\vartheta} = (\vartheta_1, \vartheta_2)$.

**5.2.7. Règle invariante de détection d'un signal à bande étroite.** Pour prouver l'existence d'une règle invariante de détection d'un signal à bande étroite noyé dans un bruit normal additif de variance inconnue, nous allons trouver la distribution simultanée des statistiques suffisantes $U_1(\mathbf{x}, \mathbf{y})$, $U_2(\mathbf{x}, \mathbf{y})$, $t(\mathbf{x}, \mathbf{y})$ (voir (5.49), (5.50)). Tout comme dans 5.2.4, normons le signal en posant $\sum_{k=1}^{N} (a_k^2 + b_k^2) = 1$ et en introduisant l'amplitude

$$\mu_N = \mu \sum_{k=1}^{N} (a_k^2 + b_k^2). \tag{5.53}$$

Comme il est facile de le voir, les statistiques $U_1(\mathbf{x}, \mathbf{y})/\sigma$, $U_2(\mathbf{x}, \mathbf{y})/\sigma$ sont indépendantes, réparties suivant les lois normales de paramètres $[(\mu_N/\sigma) \cos \varphi_0, 1]$ et $[(\mu_N/\sigma) \sin \varphi_0, 1]$ respectivement. La distribution de la statistique

$$U_3(\mathbf{x}, \mathbf{y}) = \frac{1}{\sigma^2} [U_1^2(\mathbf{x}, \mathbf{y}) + U_2^2(\mathbf{x}, \mathbf{y})] \tag{5.54}$$

est pour l'hypothèse $H$ ($\mu_N = 0$) exponentielle ($\chi^2$ à deux degrés de liberté) de densité $W_1(z) = e^{-z/2}$, $z > 0$, et pour l'alternative $K$ ($\mu_N > 0$)

$$W_1(z) = \frac{1}{2} \exp\left[-\frac{z+(\mu_N/\sigma)^2}{2}\right] I_0(\mu_N \sqrt{z}/\sigma). \tag{5.55}$$

La statistique

$$U_4(\mathrm{x}, \mathrm{y}) = \frac{1}{\sigma^2}[t(\mathrm{x}, \mathrm{y}) - U_1^2(\mathrm{x}, \mathrm{y}) - U_2^2(\mathrm{x}, \mathrm{y})] \tag{5.56}$$

est répartie suivant une loi du $\chi^2$ à $2(N-1)$ degrés de liberté.

Des raisonnements analogues à ceux utilisés pour obtenir (5.33) donnent la distribution simultanée des statistiques suffisantes considérées, soit

$$W(u_1, u_2, t \mid \vartheta_1, \vartheta_2, \vartheta_3) = C(\vartheta_1, \vartheta_2, \vartheta_3)\, h(u_1, u_2, t) \times \\ \times \exp(\vartheta_1 u_1 + \vartheta_2 u_2 - \vartheta_3 t),$$

$$-\infty < u_1 < \infty, \quad -\infty < u_2 < \infty, \quad 0 \leqslant t < \infty, \tag{5.57}$$

où

$$\vartheta_1 = \mu_N \cos\varphi_0/\sigma^2, \quad \vartheta_2 = \mu_N \sin\varphi_0/\sigma^2, \quad \vartheta_3 = 1/(2\sigma^2), \tag{5.58}$$

$$C(\vartheta_1, \vartheta_2, \vartheta_3) = [2\pi 2^{N-1}\Gamma(N-1)]^{-1}(2\vartheta_3)^N e^{-\frac{\vartheta_1^2+\vartheta_2^2}{4\vartheta_3}}, \tag{5.58'}$$

$$h(u_1, u_2, t) = (t - u_1^2 - u_2^2)^{N-2}. \tag{5.58''}$$

Introduisons le groupe de transformations

$$g(\mathbf{U}, t) = (a\mathbf{A}\mathbf{U}, a^2 t), \tag{5.59}$$

où

$$\mathbf{U} = \begin{pmatrix} U_1 \\ U_2 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} \cos\varphi_0, & -\sin\varphi_0 \\ \sin\varphi_0, & \cos\varphi_0 \end{pmatrix} \tag{5.60}$$

et $a$ une constante positive. En vertu de (5.57), on a

$$W[g(\mathbf{U}, t)] = C[\bar{g}(\boldsymbol{\vartheta}, \vartheta_3)]\, h(u_1, u_2, t) \exp(a\mathbf{A}\boldsymbol{\vartheta}\mathbf{U} - a^2\vartheta_3 t), \tag{5.61}$$

où

$$\bar{g}(\boldsymbol{\vartheta}, \vartheta_3) = (a\mathbf{A}\boldsymbol{\vartheta}, a^2\vartheta_3), \quad \boldsymbol{\vartheta} = \begin{pmatrix} \vartheta_1 \\ \vartheta_2 \end{pmatrix} \tag{5.62}$$

Comme dans le problème envisagé on a $\Theta_H = (\boldsymbol{\vartheta} = 0, \vartheta_3 > 0)$ et $\Theta_K = (|\boldsymbol{\vartheta}| \neq 0, \vartheta_3 > 0)$, il est évident que la transformation (5.62) satisfait à la condition (5.10), ce qui prouve que le problème est invariant par rapport au groupe de transformations (5.60).

On peut montrer (voir [2, 13]) que l'invariant maximal par rapport au groupe de transformations (5.59) est donné par la fonction

$$\mathrm{M}[U_1(\mathrm{x}, \mathrm{y}), U_2(\mathrm{x}, \mathrm{y}), t(\mathrm{x}, \mathrm{y})] = [U_1^2(\mathrm{x}, \mathrm{y}) + U_2^2(\mathrm{x}, \mathrm{y})]/t(\mathrm{x}, \mathrm{y}), \tag{5.63}$$

et l'invariant maximal pour le groupe (5.62) est la grandeur

$$\nu(\vartheta_1, \vartheta_2, \vartheta_3) = (\vartheta_1^2 + \vartheta_2^2)/(2\vartheta_3) = \mu_N^2/\sigma^2, \tag{5.64}$$

qui ne dépend pas de la phase initiale du signal et n'est rien d'autre que le rapport signal/bruit donné par le carré de l'amplitude du signal à la puissance moyenne du bruit.

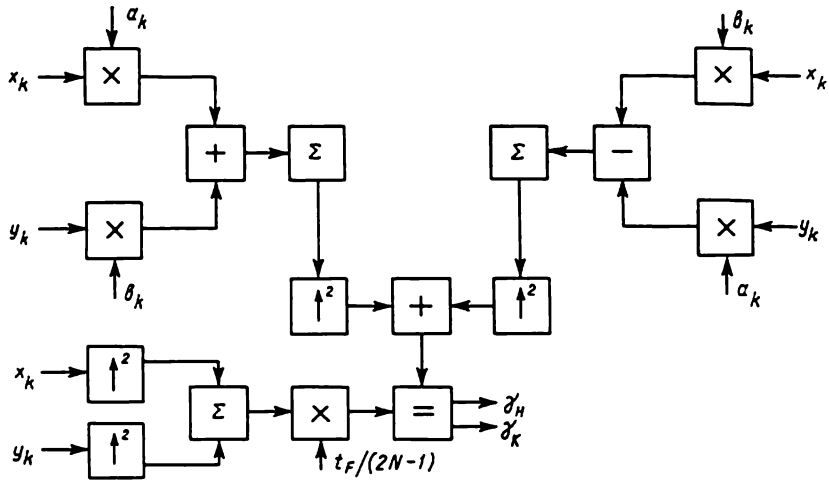



Fig. 5.3. Schéma fonctionnel du détecteur invariant d'un signal à bande étroite de phase inconnue noyé dans un bruit additif normal de variance inconnue.

Les relations (5.54), (5.55) et (5.57) permettent d'écrire la distribution simultanée des statistiques $t(\mathbf{x}, \mathbf{y})$ et $U^2(\mathbf{x}, \mathbf{y}) = U_1^2(\mathbf{x}, \mathbf{y}) + U_2^2(\mathbf{x}, \mathbf{y})$:

$$W(U^2, t \mid \vartheta, \vartheta_3) = C(\vartheta, \vartheta_3)\, h(U, t)\, e^{-\vartheta_3 t}\, \mathrm{I}_0(\vartheta U),$$

$$U \geqslant 0, \quad t \geqslant 0, \tag{5.65}$$

où $\vartheta = \sqrt{\vartheta_1^2 + \vartheta_2^2}$. Le rapport de vraisemblance pour la famille de distributions (5.65) est égal à

$$l(\mathbf{x}, \mathbf{y}) = W(U^2, t \mid \vartheta, \vartheta_3) / W(U^2, t \mid 0, \vartheta_3) =$$

$$= C_1(\vartheta, \vartheta_3)\, \mathrm{I}_0(\vartheta U), \quad U \geqslant 0, \tag{5.66}$$

et se trouve être une fonction monotone de la statistique $U^2(\mathbf{x}, \mathbf{y})$.

Ainsi, dans le cas envisagé la règle uniformément la plus puissante, sans biais et invariante, de détection d'un signal se formule comme suit: le signal est présent si

$$[U_1^2(\mathbf{x}, \mathbf{y}) + U_2^2(\mathbf{x}, \mathbf{y})] \frac{2N-1}{t(\mathbf{x}, \mathbf{y})} \geqslant C, \tag{5.67}$$

où le seuil $C$ ne dépend pas du rapport signal/bruit. La statistique dans (5.67) suit une $F$-distribution *) à $\nu_1 = 2$, $\nu_2 = 2N - 1$ degrés de liberté, centrée pour l'hypothèse $H$ et non centrée pour l'alternative $K$, de paramètre d'écart $\mu_N^2/\sigma^2$.

Le schéma fonctionnel du détecteur correspondant à l'algorithme (5.67) est donné sur la figure 5.3.

## 5.3. DÉTECTION D'UN SIGNAL NOYÉ DANS UN BRUIT NORMAL ADDITIF DE FONCTION DE CORRÉLATION INCONNUE

**5.3.1. Détection d'un signal déterministe noyé dans un bruit quasi déterministe.** Modifions le problème formulé dans 5.2.1. Pour la réalisation observée $x(t)$, il s'agit de vérifier l'hypothèse composée $H$ : $x(t) = \xi(t)$ contre l'alternative composée $K$ : $x(t) = \mu s(t) + \xi(t)$, $\mu > 0$. Pour la représentation du signal $s(t)$ et du bruit $\xi(t)$ de fonction de corrélation inconnue on utilise une même base orthogonale donnée

$$s(t) = \sum_{j=1}^{m} a_j \varphi_j(t), \tag{5.68}$$

$$\xi(t) = \sum_{j=1}^{m} \xi_j \varphi_j(t), \tag{5.69}$$

où $\mathbf{a} = (a_1, \ldots, a_m)$ est le vecteur des paramètres informationnels donnés et $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_m)$ le vecteur dont les composantes sont des variables aléatoires de valeurs moyennes nulles et de variances inconnues $\sigma_k^2 = m_1\{\xi_k^2\}$, $k = 1, \ldots, m$ (paramètres perturbateurs).

On se donne les coordonnées $x_k$ du processus $x(t)$ observé sur l'intervalle $(-T, T)$ sous la forme

$$x_k = \int_{-T}^{T} x(t)\, \varphi_k(t)\, dt, \quad k = 1, \ldots, m. \tag{5.70}$$

Il est évident que le vecteur $\mathbf{x} = (x_1, \ldots, x_m)$ suit une loi normale, de moyenne nulle pour l'hypothèse $H$, de vecteur des moyennes $\mu\mathbf{a} = (\mu a_1, \ldots, \mu a_m)$ pour l'alternative $K$, et de matrice de corrélation (tant pour l'hypothèse que pour l'alternative) dont les éléments

*) L'expression de la densité de probabilité de la $F$-distribution est donnée dans [6]. Dans cet ouvrage on peut trouver les tables des quantiles pour une loi du $F$ centrée (voir également le problème 5.2).

sont

$$m_1\{x_i x_j\} = \int_{-T}^{T}\int_{-T}^{T} \varphi_i(t)\,\varphi_j(u)\left[\sum_{k=1}^{m} \varphi_k(t)\,\varphi_k(u)\,\sigma_k^2\right] dt\,du =$$

$$= \sum_{k=1}^{m} \sigma_k^2 \delta_{ik}\delta_{jk} = \sigma_i^2 \delta_{ij}, \tag{5.71}$$

ce qui montre que les variables aléatoires $x_i$ et $x_j$ pour $i \neq j$ sont indépendantes. Il est alors facile d'écrire la distribution simultanée des variables aléatoires $x_k$, $k = 1, \ldots, m$, pour $\mu > 0$ (comparer avec (5.15)):

$$W_m(\mathbf{x}|\mu\mathbf{a},\ \boldsymbol{\sigma}^2) = (2\pi)^{-m/2} \prod_{k=1}^{m} \exp\left\{-\frac{1}{2}\sum_{k=1}^{m} (x_k - \mu a_k)^2/\sigma_k^2\right\},$$

$$\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_m^2). \tag{5.72}$$

Introduisons les désignations

$$\boldsymbol{\vartheta}' = (\vartheta_1, \ldots, \vartheta_m), \quad \vartheta_k = \mu a_k/\sigma_k^2, \tag{5.73}$$

$$\boldsymbol{\Theta}' = (\Theta_1, \ldots, \Theta_m), \quad \Theta_k = 1/(2\sigma_k^2), \tag{5.73'}$$

$$C(\boldsymbol{\vartheta}, \boldsymbol{\Theta}) = \prod_{k=1}^{m} \left(\frac{\Theta_k}{\pi}\right)^{1/2} \exp\left(-\frac{1}{4}\sum_{k=1}^{m} \vartheta_k^2/\Theta_k\right), \tag{5.74}$$

$$\mathbf{X}^2 = (x_1^2, \ldots, x_m^2); \tag{5.74'}$$

nous pouvons maintenant écrire (5.72) sous la forme

$$W_m(\mathbf{X} \mid \mu\mathbf{a},\ \boldsymbol{\sigma}^2) = C(\boldsymbol{\vartheta}, \boldsymbol{\Theta}) \exp(\boldsymbol{\vartheta}'\mathbf{X} - \boldsymbol{\Theta}'\mathbf{X}^2), \tag{5.75}$$

c'est-à-dire que (5.72) est une famille de densités exponentielles.

Le problème formulé ci-dessus de la vérification des hypothèses équivaut à celui de la vérification de l'hypothèse composée $H$: $\boldsymbol{\vartheta} = 0$, $\Theta_k > 0$, $k = 1, \ldots, m$ contre l'alternative composée $K$: $\vartheta_k = \vartheta_k^*$, $\Theta_k > 0$, $k = 1, \ldots, m$, $\boldsymbol{\vartheta}$ étant le vecteur des paramètres informationnels, et $\boldsymbol{\Theta}$ le vecteur des paramètres perturbateurs.

A partir de (5.75), on obtient l'expression du logarithme du rapport de vraisemblance, à savoir

$$\ln l(\mathbf{X} \mid \boldsymbol{\vartheta},\ \boldsymbol{\Theta}) = \ln \frac{W_m(\mathbf{X} \mid \mu\mathbf{a},\ \boldsymbol{\sigma}^2)}{W_m(\mathbf{X} \mid 0,\ \boldsymbol{\sigma}^2)} = \boldsymbol{\vartheta}'\mathbf{X} - \frac{1}{4}\sum_{k=1}^{m} \vartheta_k^2/\Theta_k. \tag{5.76}$$

Par conséquent, en vertu de 5.1.4, la règle sans biais uniformément la plus puissante de choix d'une décision pour le problème envisagé se formule comme suit: on prend la décision $\gamma_K$ que le

signal est présent si

$$\vartheta^{*\prime}\mathbf{X} = \sum_{k=1}^{m} \vartheta_k^* x_k \geqslant C(\mathbf{X}^2), \tag{5.77}$$

et la décision $\gamma_H$ qu'il n'y a pas de signal si c'est l'inégalité inverse de (5.77) qui est vérifiée.

**5.3.2. Détermination du seuil.** Ecrivons la statistique figurant dans (5.77) comme suit:

$$z = \sum_{k=1}^{m} \vartheta_k^* x_k = \sum_{k=1}^{m} \vartheta_k^* \operatorname{sgn} x_k \sqrt{x_k^2}. \tag{5.78}$$

Le vecteur $\mathbf{X}^2 = (x_1^2, \ldots, x_m^2)$ étant fixé, la variable aléatoire $z$ est discrète prenant $2^m$ valeurs pour différentes combinaisons des signes des éléments de l'échantillon $\mathbf{X}$. Dans ce cas il y a lieu d'introduire une fonction de décision randomisée (voir (5.9″))

$$\Phi(z) = \begin{cases} 1, & z > C, \\ \gamma, & z = C, \\ 0, & z < C. \end{cases} \tag{5.79}$$

Comme montré dans [14] pour une probabilité donnée $\alpha$ de fausse alarme, les constantes $\gamma$ et $C$ sont

$$\gamma = [\alpha 2^m]/2^m, \tag{5.80}$$

$$C = z_{(l)}, \quad l = [2^m(1-\alpha)], \tag{5.81}$$

où $z_{(l)}$ est la statistique d'ordre de l'échantillon $(z_1, \ldots, z_{2^m})$ obtenue à partir de l'échantillon observé $(x_1, \ldots, x_m)$, si dans (5.78) on fixe $x_k^2$, $k = 1, \ldots, m$, et qu'on envisage toutes les combinaisons possibles des signes $(\operatorname{sgn} x_1, \ldots, \operatorname{sgn} x_m)$; $[t]$ désigne le plus grand entier de t.

Pour $m \gg 1$, $[\alpha 2^m] \approx \alpha 2^m$ et alors $\gamma = 0$, donc la règle randomisée devient la règle déterministe (5.77) où le seuil $C = 2^m(1-\alpha)$.

**5.3.3. Règle invariante de détection.** Le défaut de l'algorithme (5.77) est qu'il a été obtenu pour des valeurs fixées de $\vartheta_k$ (voir (5.73)) qui, en général, sont a priori inconnues. On évite cet inconvénient en utilisant le principe d'invariance.

Introduisons le groupe de transformations $G$ des éléments de l'espace échantillonné formé par deux sous-groupes: celui de changement d'échelle $G_1$ et celui des permutations $G_2$:

$$g_1(x_k) = c_k x_k, \quad c_k > 0, \quad k = 1, \ldots, m, \tag{5.82}$$

$$g_2(\mathbf{x}) = (x_{i_1}, \ldots, x_{i_m}), \quad i_j = 1, \ldots, 2^m. \tag{5.82'}$$

En vertu de (5.75), $W_m(\mathbf{X} \mid \mu\mathbf{a}, \sigma^2)$ est invariant par rapport au groupe $G$. Il est également évident que cette transformation satisfait à la condition (5.10).

L'invariant maximal par rapport au sous-groupe $g_1$ est le vecteur des signes

$$M_1(\mathbf{x}) = (\operatorname{sgn} x_1, \ldots, \operatorname{sgn} x_m), \tag{5.83}$$

et l'invariant maximal par rapport au sous-groupe $g_2$ le vecteur des statistiques d'ordre

$$\mathbf{M}_2(\mathbf{x}) = (x^{(1)}, \ldots, x^{(m)}). \tag{5.84}$$

En vertu du théorème 2 exposé dans [11], chap. 6, l'invariant maximal par rapport au groupe $G$ est le vecteur

$$\mathbf{M}(\mathbf{x}) = (\operatorname{sgn} x^{(1)}, \ldots, \operatorname{sgn} x^{(m)}), \tag{5.85}$$

ou la statistique

$$\mathbf{M}(\mathbf{x}) = \frac{1}{2} \sum_{i=1}^{m} (1 + \operatorname{sgn} x_i), \tag{5.86}$$

égale au nombre de valeurs positives dans l'échantillon $\mathbf{X}$.

Ainsi, l'algorithme invariant de détection d'un signal dans le cas envisagé est l'algorithme de signe, unilatéral si toutes les composantes du vecteur $\mathbf{a}$ sont positives, et bilatéral si ces composantes ont des signes arbitraires. Les seuils pour une probabilité donnée de fausse alarme sont donnés par des distributions binomiales (voir 3.2.1, 3.2.2).

**5.3.4. Détection d'un signal harmonique dans un bruit normal à paramètres inconnus du spectre énergétique.** Nous allons envisager le problème de la détection d'un signal harmonique $\mu s(t) = \mu \cos \omega_0 t$ de fréquence donnée $\omega_0$ et d'amplitude inconnue $\mu > 0$ noyé dans un bruit normal additif stationnaire dont les valeurs échantillonnées $x_i = x(t_i)$, $i = 1, 2, \ldots, n$, satisfont à l'équation d'autorégression d'ordre $k$ (voir (4.159))

$$\sum_{j=0}^{k} \lambda_j x_{i-j} = \varepsilon_i, \quad i = k+1, k+2, \ldots,$$

où $\lambda_j$ sont les coefficients inconnus de la régression (paramètres perturbateurs) et $\varepsilon_i$ une suite de variables aléatoires normales indépendantes de valeurs moyennes nulles et de variances unités. Le spectre énergétique de ce bruit est (voir [15, 17]).

$$F(\omega) = \left[2\pi \left(F_0 + 2 \sum_{l=1}^{k} F_l \cos l\omega\tau\right)\right]^{-1}, \tag{5.87}$$

où $\tau = t_{i+1} - t_i$, $i = 1, 2, \ldots$ et

$$F_l = \sum_{i=l}^{k} \lambda_i \lambda_{i-l}. \tag{5.87'}$$

Les éléments de la matrice $\mathbf{C}\, n \times n$, inverse de la matrice de corrélation du bruit, sont

$$c_{ij} = F_{|i-j|}, \quad |i-j| \leqslant k < i, \quad j \leqslant n-k, \tag{5.88}$$

$$c_{n-i+1,\, n-j+1} = F_{|i-j|} - \sum_{l=|i-j|+i}^{k} \lambda_l \lambda_{l+|i-j|}, \quad |i-j| \leqslant k, \; i, \; j \leqslant k. \tag{5.88'}$$

Dans le cas envisagé la loi de distribution de l'échantillon $\mathbf{x} = (x_1, \ldots, x_n)$ de la somme du signal et du bruit peut être approchée par une famille exponentielle

$$w(\mathbf{x} \mid \Theta, \boldsymbol{\vartheta}) = C(\Theta, \boldsymbol{\vartheta}) \exp\{\Theta U(\mathbf{x}) + \boldsymbol{\vartheta}' \mathbf{r}(\mathbf{x})\} \tag{5.89}$$

de paramètres

$$\Theta = \mu\left(F_0 + 2\sum_{l=1}^{k} F_l \cos l\omega_0\tau\right), \tag{5.90}$$

$$\boldsymbol{\vartheta} = (\vartheta_0, \ldots, \vartheta_k), \; \vartheta_0 = -F_0/2, \; \vartheta_j = -F_j, \; j = 1, \ldots, k, \tag{5.90'}$$

et de statistiques

$$U(\mathbf{x}) = \mathbf{x}'\mathbf{s} = \sum_{i=1}^{n} x_i \cos \omega_0 t_i, \tag{5.91}$$

$$\mathbf{r}(\mathbf{x}) = [r_1(\mathbf{x}), \ldots, r_k(\mathbf{x})], \quad r_l(\mathbf{x}) = \sum_{i=1}^{n} x_i x_{i-l}, \tag{5.91'}$$

l'approximation étant d'autant meilleure que la taille $n$ de l'échantillon est plus grande.

Dans cette approximation, le problème envisagé de la détection d'un signal harmonique se formule comme suit : vérifier l'hypothèse $H$ (il n'y a pas de signal) $\Theta = 0$, $\boldsymbol{\vartheta}$ étant inconnu, contre l'alternative $K$ (le signal est présent) : $\Theta > 0$, $\boldsymbol{\vartheta}$ étant inconnu.

Si l'on arrivait à trouver une statistique $Y(\mathbf{x}) = Y(U(\mathbf{x}), \mathbf{r}(\mathbf{x}))$, qui

a) soit monotone par rapport à $U$ pour chaque $\mathbf{r}$,

b) ne dépende pas de $\mathbf{r}$ pour l'hypothèse $H$, la règle uniformément la plus puissante de détection d'un signal pourrait alors se formuler comme suit : le signal est présent (l'hypothèse $H$ est rejetée) si

$$Y(\mathbf{x}) \geqslant C, \tag{5.92}$$

où le seuil $C$ se trouve d'après une probabilité donnée de fausse alarme.

La recherche de la statistique mentionnée pour des échantillons de taille $n$ finie se heurte à de grosses difficultés. Il est néanmoins possible de donner une statistique satisfaisant aux conditions a) et b) lorsque $n \to \infty$. Pour le cas envisagé une telle statistique peut s'écrire comme suit [10]:

$$Y(\mathbf{x}) = U(\mathbf{x}) \left(\hat{F}_0 + 2 \sum_{l=1}^{k} \hat{F}_l \cos l\omega_0\tau\right)^{1/2} \Big/ \left(\sum_{j=1}^{n} \cos^2 \omega_0 t_j\right)^{1/2}, \quad (5.93)$$

où

$$\hat{F}_l = \sum_{l=1}^{k} \hat{\lambda}_l \hat{\lambda}_{i-l}, \quad (5.93')$$

$\hat{\lambda}_l(\mathbf{x})$ étant les valeurs asymptotiques (pour $n \to \infty$) des estimations du maximum de vraisemblance du coefficient de régression (voir 4.4.7)

## 5.4. ALGORITHMES DE CLASSIFICATION SEMBLABLES ET INVARIANTS

**5.4.1. Position du problème.** Dans le chapitre précédent (voir 4.1) nous avons examiné les algorithmes adaptatifs de classification des observations, obtenus par substitution dans le rapport de vraisemblance des estimations du maximum de vraisemblance des moyennes inconnues ou des matrices des covariances, estimations obtenues à partir des échantillons d'apprentissage classifiés. Ces algorithmes étaient consistants, convergeant en probabilité vers les algorithmes applicables dans les conditions où l'on dispose d'une information a priori complète, et lorsque la taille des échantillons d'apprentissage augmente indéfiniment. Il est intéressant d'obtenir les algorithmes de classification adaptatifs, qui soient doués de certaines propriétés d'optimalité pour des tailles finies des échantillons d'apprentissage.

Considérons deux classes $S_1$ et $S_2$ caractérisées par les densités $W_N(\mathbf{x}; \boldsymbol{\vartheta}_1)$ et $W_N(\mathbf{x}; \boldsymbol{\vartheta}_2)$ $N$-dimensionnelles où les paramètres *inconnus* $\boldsymbol{\vartheta}_1$ et $\boldsymbol{\vartheta}_2$ sont les points de l'espace euclidien $\Theta$ à $m$ dimensions. On a deux échantillons d'apprentissage classifiés $\mathbf{x}_1^{(1)}, \ldots, \mathbf{x}_{n_1}^{(2)}$ et $\mathbf{x}_1^{(2)}, \ldots, \mathbf{x}_{n_2}^{(2)}$ appartenant aux classes $S_1$ et $S_2$ respectivement. Le problème de la classification consiste à rapporter l'échantillon observé (informationnel) $\mathbf{x}_1, \ldots, \mathbf{x}_n$ à l'une des classes, $S_1$ ou $S_2$. En termes de la vérification des hypothèses statistiques, ce problème à trois échantillons peut être formulé comme un problème de vérification des hypothèses composées sur les paramètres de trois

échantillons (dont deux sont d'apprentissage et un informationnel):

$$H: (\vartheta_1, \vartheta_2, \vartheta_1), \ \vartheta_i \in \Theta, \ i = 1; 2,$$
$$K: (\vartheta_1, \vartheta_2, \vartheta_2), \ \vartheta_i \in \Theta, \ i = 1; 2.$$

Ecrivons l'algorithme de classification en utilisant la fonction de décision non randomisée $\Phi(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x})$, où $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$, $\mathbf{x}$ sont les vecteurs colonnes des échantillons d'apprentissage et informationnel. On a alors

$$\Phi(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}) = \begin{cases} 1, & \mathbf{x} \in S_2, \\ 0, & \mathbf{x} \in S_1. \end{cases} \tag{5.94}$$

A l'aide de (5.94) il est facile d'écrire les probabilités de classification erronée. La probabilité $\alpha$ de classification erronée de première espèce est

$$\alpha = P\{\Phi = 1 \mid H\} = m_1\{\Phi(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}) \mid H\}. \tag{5.95}$$

La probabilité $\beta$ de classification erronée de seconde espèce est

$$\beta = P\{\Phi = 0 \mid K\} = 1 - P\{\Phi = 1 \mid K\} = 1 - m_1\{\Phi(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}) \mid K\}. \tag{5.96}$$

Les paramètres $\vartheta_1$, $\vartheta_2$ étant inconnus, il est naturel d'exiger que l'une des probabilités d'erreur (par exemple, $\alpha$) n'en dépende pas, et l'autre soit la plus petite possible dans la classe de tous les algorithmes de classification. L'algorithme uniformément le plus puissant, dans la classe des algorithmes semblables, satisfait à cette condition. Cependant, comme nous l'avons noté à ce propos, les algorithmes semblables uniformément les plus puissants ou bien n'existent pas, ou bien leur construction se heurte à de nombreuses difficultés. On peut essayer de résoudre le problème en introduisant certaines limitations supplémentaires correspondant aux conditions d'application pratique des algorithmes de classification. On peut exiger, par exemple, que les algorithmes cherchés soient invariants par rapport à certains groupes de transformations de l'espace échantillonné (5.1.5).

**5.4.2. Algorithme optimal de classification de variables aléatoires normales unidimensionnelles.** Commençons par le cas le plus simple envisagé dans 4.1.2. Supposons que les classes $S_1$ et $S_2$ soient caractérisées par des distributions normales de moyennes inconnues $a_1$ et $a_2$ et de variances données $\sigma^2$. Par apprentissage avec maître, on a obtenu les échantillons classifiés indépendants: $x_1^{(1)}, \ldots, x_{n_1}^{(1)}$ de la classe $S_1$ et $x_1^{(2)}, \ldots, x_{n_2}^{(2)}$ de la classe $S_2$. Proposons-nous de trouver une règle optimale de classification, qui permette de rapporter l'échantillon (informationnel) observé $x_1, \ldots, x_n$, indépendant des échantillons d'apprentissage, à l'une ou l'autre classe.

On sait que la moyenne échantillonnée est une statistique suffisante pour la moyenne inconnue d'une distribution normale dont on donne la variance. Désignons les moyennes échantillonnées comme suit :

$$\hat{a}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i^{(1)}, \quad \hat{a}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} x_i^{(2)}; \tag{5.97}$$

$$\hat{a} = \frac{1}{n} \sum_{i=1}^{n} x_i. \tag{5.97'}$$

Il est facile de voir que la transformation linéaire

$$t_1 (x^{(1)}, x) = \frac{1}{n_1 + n} (n_1 \hat{a}_1 + n\hat{a}), \tag{5.98}$$

$$t_2 (x^{(2)}) = \hat{a}_2, \tag{5.98'}$$

$$t_3 (x^{(1)}, x) = \hat{a} - \hat{a}_1 \tag{5.98''}$$

permet de formuler le problème envisagé en termes de la vérification des hypothèses statistiques sur les valeurs moyennes des statistiques $t_1$, $t_2$, $t_3$ comme suit : on vérifie l'hypothèse composée $H$ : $(a_1, a_2, 0)$ contre l'alternative composée $K$ : $\left(\frac{n_1 a_1 + n a_2}{n_1 + n}, a_2, a_2 - a_1\right)$.

Introduisons la statistique vectorielle $\mathbf{t}(x^{(1)}, x^{(2)}, x) = (t_1, t_2)$. Il est évident que la moyenne conditionnelle $m_1 \{\Phi (x^{(1)}, x^{(2)}, x) \mid H, \mathbf{t}\}$ ne dépend pas des paramètres $a_1$, $a_2$. Ensuite, la famille des distributions normales $W(\mathbf{t} \mid a_1, a_2)$ étant complète, dans le cas envisagé il existe un algorithme semblable de classification de structure de Neyman (voir 5.1.4). Comme le rapport de vraisemblance est égal à

$$l(t_1, t_2, t_3) = C(a_1, a_2) \exp \left\{ \frac{n(a_2 - a_1)}{(n_1 + n)\sigma_1^2} t_1 + \frac{a_2 - a_1}{\sigma_3^2} t_3 \right\}, \tag{5.99}$$

où $\sigma_1^2 = \sigma^2/(n_1 + n)$, $\sigma_3^2 = \sigma^2 (1/n_1 + 1/n)$, le rapport de vraisemblance conditionnel, pour $\mathbf{t} = (t_1, t_2)$ fixé, ne dépend que de la statistique $t_3$. Par conséquent, il doit exister une règle de classification sans biais uniformément la plus puissante, revenant à comparer $|t_3|$ avec un seuil $C(t)$, qui pour $\alpha$ donné s'obtient à partir de la relation

$$m_1 \{\Phi (\mathbf{t}, t_3) \mid \mathbf{t}, H\} = \alpha$$

pour tous les $\mathbf{t}$.

Il est facile de voir que pour l'hypothèse $H$ la statistique $t_3$ ne dépend pas de $t_2$, ni de $t_1$ si $n_1 = n$. Dans ce cas le seuil ne dépend pas de $\mathbf{t}$ et la règle optimale de classification peut se formuler comme

suit : l'observation $\mathbf{x}$ se rapporte à la classe $S_2$ si

$$|t_3| = |\hat{a} - \hat{a}_1| \geqslant \sqrt{\frac{2}{n}}\,\sigma x_\alpha, \tag{5.100}$$

et se rapporte à la classe $S_1$ si c'est l'inégalité inverse de (5.100) qui est vérifiée ; $x_\alpha$ est ici l'$\alpha$-quantile de la loi normale normalisée.

Remarquons que la règle (5.100) utilise un seul échantillon d'apprentissage. Autrement dit, l'apprentissage préalable dans le cas envisagé n'est indispensable que pour l'une des classes.

**5.4.3. Algorithme de classification invariant uniformément le plus puissant.** Pour les conditions du problème envisagé dans le paragraphe précédent, on examine dans [4] l'algorithme de classification uniformément le plus puissant, invariant par rapport aux deux groupes de transformations : le groupe des translations $G_1$ et le groupe des identités $G_2$ avec éventuellement le changement de signe. L'invariance par rapport à $G_1$ implique de considérer deux statistiques seulement, à savoir

$$U_1(\mathbf{x}^{(1)}, \mathbf{x}) = \frac{\hat{a}_1 - \hat{a}}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n}}}, \quad U_2(\mathbf{x}^{(2)}, \mathbf{x}) = \frac{\hat{a}_2 - \hat{a}}{\sigma\sqrt{\frac{1}{n_2} + \frac{1}{n}}}. \tag{5.101}$$

Les statistiques $U_1$ et $U_2$ suivent une loi normale bidimensionnelle dont les paramètres sont : pour l'hypothèse $H$

$$m_1\{U_1|H\} = 0, \quad m_1\{U_2|H\} = \frac{a_2 - a_1}{\sigma\sqrt{\frac{1}{n_2} + \frac{1}{n}}}, \tag{5.102}$$

$$M_2\{U_1|H\} = M_2\{U_2|H\} = 1, \tag{5.102'}$$

$$\rho = m_1\{U_1U_2|H\} = \frac{1}{n}\left[\left(\frac{1}{n_1} + \frac{1}{n}\right)\left(\frac{1}{n_2} + \frac{1}{n}\right)\right]^{-1/2}, \tag{5.102''}$$

et pour l'alternative $K$

$$m_1\{U_1|K\} = \frac{a_1 - a_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n}}}, \quad m_1\{U_2|K\} = 0, \tag{5.103}$$

$$M_2\{U_1|K\} = M_2\{U_2|K\} = 1, \quad m_1\{U_1U_2|K\} = \rho. \tag{5.103'}$$

Les statistiques $U_1$ et $U_2$ peuvent être décorrélées par une transformation linéaire

$$V(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}) = \frac{1}{\sqrt{1-\rho^2}}\,[\rho U_1(\mathbf{x}^{(1)}, \mathbf{x}) - U_2(\mathbf{x}^{(2)}, \mathbf{x})], \tag{5.104}$$

$$U(\mathbf{x}^{(1)}, \mathbf{x}) \equiv U_1(\mathbf{x}^{(1)}, \mathbf{x}). \tag{5.104'}$$

Les statistiques $V$ et $U$ sont indépendantes, normales, de variance unité et de moyennes égales à

$$m_1\{V|H\}=\frac{a_1-a_2}{\sigma}\left(\frac{1}{n_2}+\frac{1}{n}\right)^{-1/2}, \quad m_1\{U|H\}=0, \tag{5.105}$$

$$m_1\{V|K\}=\frac{\rho}{\sqrt{1-\rho^2}}\,\frac{a_2-a_1}{\sigma}\left(\frac{1}{n_1}+\frac{1}{n}\right)^{-1/2}, \tag{5.105'}$$

$$m_1\{U|K\}=\frac{a_2-a_1}{\sigma}\left(\frac{1}{n_2}+\frac{1}{n_1}\right)^{-1/2}. \tag{5.105''}$$

Il est montré dans [4] que l'existence d'un tel algorithme de classification utilisant les statistiques $V$ et $U$ implique l'invariance par rapport à la transformation $G_2$. La règle de classification uniformément la plus puissante obtenue compte tenu de cette restriction se formule comme suit: pour une probabilité donnée $\alpha$ d'erreur de première espèce, on adopte l'hypothèse $H$ si

$$-R\,[V(\mathbf{x}^{(1)},\ \mathbf{x}^{(2)},\ \mathbf{x})]<U(\mathbf{x}^{(2)},\ \mathbf{x})<R\,[V(\mathbf{x}^{(1)},\ \mathbf{x}^{(2)},\ \mathbf{x})], \tag{5.106}$$

où $y = R(x)$ est la fonction inverse de la fonction

$$x=\frac{\sqrt{1-\rho^2}}{2\rho}\{F^{-1}[2-\alpha-F(y)]-y\}, \tag{5.106'}$$

et $F(y)$ l'intégrale de Laplace. L'hypothèse $H$ est rejetée si l'une des inégalités (5.106) ne se trouve pas vérifiée.

En comparant (5.106) et (4.9) on peut remarquer *) que l'algorithme de classification invariant uniformément le plus puissant diffère notablement de l'algorithme de classification obtenu par substitution, dans le rapport de vraisemblance, des estimations du maximum de vraisemblance des moyennes inconnues.

**5.4.4. Classification des ensembles normaux multidimensionnels.** Nous allons envisager le problème de la classification dans le cas de la symétrie sphérique. Les classes $S_1$ et $S_2$ appartiennent à des ensembles normaux à $N$ dimensions, caractérisés par des vecteurs des moyennes inconnus $\mathbf{a}_1$ et $\mathbf{a}_2$ respectivement et les matrices de corrélation données égales à $\sigma^2\mathbf{I}$. D'une manière analogue à ce qui a été exposé dans 5.4.2, on peut montrer que pour le cas multidimensionnel une règle semblable à structure de Neyman utilise la statistique

$$\mathbf{t}_3(\mathbf{x}^{(1)},\ \mathbf{x})=\hat{\mathbf{a}}-\hat{\mathbf{a}}_1, \tag{5.107}$$

*) Il est évident que pour la classification d'un échantillon informationnel $(x_1, \ldots, x_n)$ de dimension $n > 1$ il y a lieu dans (4.9) de remplacer $x$ par $\hat{a} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

où $\hat{\mathbf{a}}$, $\hat{\mathbf{a}}_1$ sont les estimations vectorielles, égales à

$$\hat{\mathbf{a}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i, \quad \hat{\mathbf{a}}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{x}_i^{(1)}. \tag{5.107'}$$

Cependant dans le cas vectoriel il est bien plus difficile d'écrire, sous forme explicite, l'algorithme non biaisé uniformément le plus puissant que dans le cas scalaire. Il y a lieu, à cet effet, de se limiter à la classe des algorithmes, invariants par rapport au groupe de transformations orthogonales de l'espace des échantillons à $N$ dimensions. L'invariant maximal pour ce groupe de transformation est donné par la fonction

$$\mathrm{M}(\mathbf{x}^{(1)}, \mathbf{x}) = \frac{\mathbf{t}_3'\mathbf{t}_3}{\sigma^2(1/n_1 + 1/n)} = \frac{\sum_{k=1}^{N} t_{3k}^2}{\sigma^2(1/n_1 + 1/n)}. \tag{5.108}$$

Comme la statistique $\mathbf{t}_3$ suit une loi normale à $N$ dimensions de matrice de corrélation égale à $\sigma^2\left(\frac{1}{n_1} + \frac{1}{n}\right)\mathbf{I}$ et de vecteur des moyennes égal à zéro pour l'hypothèse $H$ et à $\mathbf{a}_2 - \mathbf{a}_1$ pour l'alternative $K$, l'invariant maximal (5.108) suit pour l'hypothèse $H$ la loi du $\chi^2$ centrée à $N$ degrés de liberté, et pour l'alternative $K$ la loi du $\chi^2$ non centrée avec le même nombre de degrés de liberté et le paramètre d'écart égal à

$$\delta = \frac{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1}}{\sigma^2} (\mathbf{a}_2 - \mathbf{a}_1)'(\mathbf{a}_2 - \mathbf{a}_1). \tag{5.109}$$

Mais le rapport de vraisemblance

$$l(\mathrm{M}) = \frac{W_{\chi^2}(\mathrm{M}, \delta)}{W_{\chi^2}(\mathrm{M}, 0)} \tag{5.110}$$

est une fonction monotone de M, par conséquent la règle uniformément la plus puissante de classification dans le cas envisagé (voir 5.1.5) se formule comme suit : pour une probabilité donnée $\alpha$ d'erreur de première espèce, on adopte l'alternative $K$ si

$$\mathrm{M}(\mathbf{x}^{(1)}, \mathbf{x}) \geqslant \chi_\alpha^2, \tag{5.111}$$

et l'hypothèse $H$ si c'est l'inégalité inverse de (5.111) qui est vérifiée. Dans la formule (5.111) la quantité $\chi_\alpha^2$ est l'$\alpha$-quantile de la loi du $\chi^2$ centrée à $N$ degrés de liberté.

## 5.5. PROBLÈMES

**5.1.** On dispose d'un échantillon indépendant de taille $n$ appartenant à la distribution uniforme $w_1(x) = 1/b$, $a \leqslant x \leqslant a + b$. Montrer que la règle semblable uniformément la plus puissante de vérification de l'hypothèse $H$: $a = 0$ contre l'alternative composée $K$ : $a \neq 0$ est basée sur la comparaison à un seuil de la statistique

$$y_n(\mathbf{x}) = x^{(1)}/(x^{(n)} - x^{(1)}), \tag{1}$$

où $x^{(1)}$ et $x^{(n)}$ sont les statistiques d'ordre, minimale et maximale.

**5.2.** [6]. Soient deux échantillons indépendants $x_1, \ldots, x_{n_1}$ et $y_1, \ldots, y_{n_2}$ d'une loi normale de moyennes et de variances inconnues. Il y a lieu de vérifier l'hypothèse $H$ contre l'alternative $K$ que $\sigma_x^2 > \sigma_y^2$. Montrer que l'algorithme semblable uniformément le plus puissant de vérification de l'hypothèse est de la forme (pour une probabilité donnée $\alpha$ d'erreur de première espèce)

$$s_1^2/s_2^2 \geqslant F_\alpha(n_1 - 1, \quad n_2 - 1), \tag{2}$$

où

$$s_k^2 = \frac{1}{n_k - 1}\left[\sum_{i=1}^{n_k}\left(x_i - \frac{1}{n_k}\sum_{j=1}^{n_k} x_j\right)^2\right], \quad k = 1;\ 2, \tag{3}$$

et $F_\alpha(n_1 - 1, n_2 - 1)$ est le quantile de la distribution $F$ à $n_1 - 1$, $n_2 - 1$ degrés de liberté. La densité de probabilité de la distribution $F$ à $n - 1$, $m - 1$ degrés de liberté est donnée par la formule

$$w_1(z) = \Gamma[(m+n-2)/2]\{\Gamma[(m-1)/2]\,\Gamma[(n-1)/2]\}^{-1} \times$$
$$\times \left(\frac{n-1}{m-1}\right)^{\frac{n-1}{2}} z^{\frac{n-1}{2}-1}\left(1 + \frac{n-1}{m-1}z\right)^{-(m+n-2)/2}. \tag{4}$$

Vérifier que la probabilité d'erreur de seconde espèce dépend seulement de $\sigma_x/\sigma_y$.

**5.3.** (Test de Stein [11].) Considérons le problème de la détection d'un signal constant $s(t \equiv \mu$ (amplitude inconnue) noyé dans un bruit normal additif de moyenne nulle et de variance inconnue. Supposons que l'on prélève un échantillon indépendant $x_1, \ldots, x_{n_0}$ de taille $n_0$ à partir duquel on calcule l'estimation sans biais de la variance du bruit:

$$s^2 = \frac{1}{n_0 - 1}\left[\sum_{i=1}^{n_0}\left(x_i - \frac{1}{n_0}\sum_{j=1}^{n_0} x_j\right)^2\right]. \tag{5}$$

puis on fait $n - n_0$ observations indépendantes supplémentaires, avec

$$n = \max\left\{\left[\frac{s^2}{c}\right] + 1, \quad n_0 + 1\right\}, \tag{6}$$

où $c_0 > 0$ est une constante donnée et $[z]$ le plus grand entier $\leqslant z$. Montrer que la statistique

$$y_n(\mathbf{x}) = \sum_{i=1}^{n} a_i x_i / \sqrt{c_0}, \tag{7}$$

où les coefficients de pondération satisfont à la condition

$$a_1 = \ldots = a_{n_0}, \quad a_{n_0+1} = \ldots = a_n, \quad \sum_{i=1}^{n} a_i = 1,$$

$$s^2 \sum_{i=1}^{n} a_i^2 = c_0, \tag{8}$$

suit une loi centrée de Student à $n_0 - 1$ degrés de liberté.

Vérifier que pour les algorithmes de détection tels que la décision sur la présence du signal est prise si pour $\mu > 0$ on a

$$y_n(\mathbf{x}) \geqslant t_\alpha(n_0 - 1), \tag{9}$$

et si pour un signe quelconque

$$|y_n(\mathbf{x})| \geqslant t_{\alpha/2}(n_0 - 1), \tag{10}$$

où $t_\lambda(m)$ est l'$\lambda$-quantile de la loi de Student à $m$ degrés de liberté, la probabilité de détection correcte ne dépend pas de la variance du bruit et se trouve être une fonction croissante de $|\mu|$. Dans (9) et (10), $\alpha$ est la probabilité donnée de fausse alarme.

5.4. Montrer que la règle sans biais uniformément la plus puissante, basée sur l'observation de la réalisation $x(t)$, $|t| \leqslant T$, de détection du signal $\mu s(t)$, où $\mu$ est l'amplitude inconnue du signal et $s(t)$ une fonction donnée, noyé dans un bruit additif normal de moyenne nulle et de fonction de corrélation donnée $B(t, y)$, peut s'écrire comme suit : on prend la décision que le signal est présent si

$$\left| \int_{-T}^{T} V(t)\, x(t)\, dt \right| \geqslant d_T x_{\alpha/2}, \tag{11}$$

où $V(t)$ est solution de l'équation intégrale,

$$\int_{-T}^{T} B(t, y)\, V(y)\, dy = s(t), \quad |t| \leqslant T, \tag{12}$$

$x_{\alpha/2}$ est l'$\alpha/2$-quantile de la loi normale correspondant à une probabilité donnée de fausse alarme, et

$$d_T = \int_{-T}^{T} V(t)\, s(t)\, dt. \tag{13}$$

5.5. Pour les conditions du problème 5.4, envisager le cas où on sait a priori que $\mu > 0$ et montrer que dans ce cas la règle uniformément la plus puissante consiste à adopter l'hypothèse sur la présence du signal si

$$\int_{-T}^{T} V(t)\, x(t)\, dt \geqslant d_T x_\alpha. \tag{14}$$

5.6. Soient deux signaux $\mu_1 s_1(t)$ et $\mu_2 s_2(t)$, $\mu_1$, $\mu_2$ étant les amplitudes inconnues et $s_1(t)$, $s_2(t)$ des fonctions données, noyés dans un bruit additif normal de moyenne nulle et de fonction de corrélation $B(t, y)$ donnée. Montrer que la règle de discrimination de ces signaux, optimale d'après le critère du maximum de vraisemblance et basée sur l'observation de la réalisation $x(t)$, $|t| \leqslant T$, peut s'écrire comme suit : on décide qu'on est en présence du signal $\mu_1 s_1(t)$ si

$$\frac{1}{d_{T_1}} \int_{-T}^{T} V_1(t)\, x(t)\, dt \geqslant \frac{1}{d_{T_2}} \int_{-T}^{T} V_2(t)\, x(t)\, dt, \tag{15}$$

où $V_1(t)$ et $V_2(t)$ sont solutions des équations intégrales

$$\int_{-T}^{T} B(t, y)\, V_1(y)\, dy = s_1(t), \quad |t| \leqslant T, \tag{16}$$

$$\int_{-T}^{T} B(t, y)\, V_2(y)\, dy = s_2(t), \quad |t| \leqslant T \tag{17}$$

et

$$d_{T_i} = \int_{-T}^{T} V_i(t)\, s_i(t)\, dt, \quad i = 1\,;\ 2. \tag{18}$$

# ANNEXES

## *ANNEXE 1*

## ESTIMATION D'UNE DENSITÉ DE PROBABILITÉ INCONNUE

**A.1.1. Méthode des fonctions potentielles.** Si un échantillon indépendant $\mathbf{y} = (y_1, \ldots, y_M)$ appartient à la distribution $F_1(x)$, la fonction de répartition empirique

$$F_1^*(x, \mathbf{y}) = \frac{1}{M} \sum_{i=1}^{M} u(x - y_i) \tag{1}$$

converge en probabilité vers $F_1(x)$ pour $M \to \infty$, donc représente une estimation consistante de $F_1(x)$. En dérivant d'une manière formelle les deux membres de la formule (1) on obtient l'estimation de la densité de probabilité

$$w_1^*(x, \mathbf{y}) = \frac{1}{M} \sum_{i=1}^{M} \delta(x - y_i). \tag{2}$$

La fonction (2) devient infinie pour $x = y_i$, $i = 1, \ldots, M$, et est égale à zéro si l'argument $x$ ne coïncide avec aucune valeur échantillonnée. Il est désirable que l'estimation de la densité soit obtenue à partir de (2) en remplaçant $\delta(x - y_i)$ par des fonctions « lissées ». Ces estimations peuvent être obtenues par la *méthode des fonctions potentielles.*

Supposons que dans l'espace fonctionnel linéaire on ait $m$ fonctions linéairement indépendantes (bases)

$$\mathbf{\Phi}_m(x) = [\varphi_1(x), \ldots, \varphi_m(x)], \tag{3}$$

la matrice $\mathbf{a}$ formée par les produits scalaires de ces fonctions

$$\int_{-\infty}^{\infty} \varphi_k(x)\, \varphi_l(x)\, dx = a_{kl} < \infty \tag{4}$$

n'étant pas dégénérée. Définissons comme suit la fonction delta lissée :

$$\delta_m(x, z) = \mathbf{\Phi}_m'(x)\, \mathbf{a}^{-1} \mathbf{\Phi}_m(z). \tag{5}$$

Pour une fonction quelconque $f(x)$ représentable comme

$$f(x) = \Phi'_m(x)\,\mathbf{c} = \sum_{i=1}^{m} c_i \varphi_i(x), \tag{6}$$

on a

$$\int_{-\infty}^{\infty} \delta_m(x, z) f(z)\, dz = \Phi'_m(x)\,\mathbf{c}\mathbf{a}^{-1} \int_{-\infty}^{\infty} \Phi_m(z)\, \Phi'_m(z)\, dz = \Phi'_m(x)\,\mathbf{c},$$

c'est-à-dire que l'on obtient une relation analogue à celle que l'on connaît pour la fonction delta (effet de filtrage):

$$\int_{-\infty}^{\infty} \delta_m(x, z) f(z)\, dz = f(x). \tag{7}$$

Si $\Phi_m$ est une base orthogonale, on a $a_{kl} = a_k$ pour $k = l$, $a_{kl} = 0$ pour $k \neq l$, et à partir de (5)

$$\delta_m(x, z) = \sum_{k=1}^{m} \frac{1}{a_k} \varphi_k(x)\, \varphi_k(z). \tag{8}$$

M. Aïzerman, D. Braverman et L. Rozonoer [1] ont proposé d'appeler la fonction delta (8) lissée *fonction potentielle*.

Pour une base orthonormée on a $a_k = 1$, $k = 1, \ldots, m$, et alors

$$\delta_m(x, z) = \Phi'_m(x)\, \Phi_m(z) = \sum_{k=1}^{m} \varphi_k(x)\, \varphi_k(z)\,; \tag{9}$$

dans ce cas le système de fonctions $\Phi_m$ *ne doit pas être complet*, sinon $\delta_m(x, z)$ devient de nouveau une fonction delta.

Pour une base orthonormée l'estimation de la densité de probabilité peut s'écrire comme (comparer avec (2))

$$\hat{w}_1(x, \mathrm{y}) = \frac{1}{M} \sum_{i=1}^{M} \delta_m(x, y_i) = \frac{1}{M} \sum_{i=1}^{M} \sum_{k=1}^{m} \varphi_k(x)\, \varphi_k(y_i). \tag{10}$$

L'estimation (10) converge vers $w_1(x)$ en moyenne quadratique, c'est-à-dire

$$\lim_{M \to \infty} \int_{\mathrm{Y}} [w_1(x) - \hat{w}_1(x, \mathrm{y})]^2\, W(\mathrm{y})\, d\mathrm{y} = 0$$

uniformément pour toutes les valeurs de $x$.

L'algorithme (10) d'estimation de la densité peut s'écrire sous forme récurrente. Nous allons écrire (10) comme suit:

$$\hat{w}_1(x, \mathbf{y}) = \frac{1}{M}\delta_m(x, y_M) + \frac{1}{M-1}\sum_{i=1}^{M-1}\delta_m(x-y_i) -$$

$$-\frac{1}{M(M-1)}\sum_{i=1}^{M-1}\delta_m(x, y_i). \qquad (11)$$

$M$ étant un entier positif quelconque, en introduisant la désignation

$$\hat{w}_{1r}(x, \mathbf{y}) = \frac{1}{r}\sum_{i=1}^{r}\delta(x, y_i), \quad r = 1, \ldots, M, \qquad (12)$$

on obtient à partir de (11) l'algorithme de récurrence suivant:

$$\hat{w}_{1r}(x, \mathbf{y}_1^r) = \hat{w}_{1,r-1}(x, \mathbf{y}_1^{r-1}) + \frac{1}{r}[\delta_m(x, y_r) - \hat{w}_{1,r-1}(x, \mathbf{y}_1^{r-1})], \qquad (13)$$

où $\mathbf{y}_1^r = (y_1, \ldots, y_r)$. Il est facile de voir que la procédure d'estimation de la densité de probabilité est, en vertu de (12), un cas particulier de l'algorithme de l'*approximation stochastique* (voir annexe 2).

La méthode envisagée peut être généralisée à l'estimation d'une densité de probabilité multidimensionnelle $w_n(\mathbf{x}) = w_n(x_1, \ldots, x_n)$. Soit $\mathbf{y}_i = (y_{i1}, \ldots, y_{iM})$ une suite de vecteurs appartenant à la distribution $w_n(\mathbf{x})$ et soit

$$\mathbf{\Phi}_m(\mathbf{x}) = [\varphi_1(\mathbf{x}), \ldots, \varphi_m(\mathbf{x})], \qquad (14$$

$$\varphi_l(\mathbf{x}) = \varphi_l(x_1, \ldots, x_n),$$

$$\int_X \varphi_l(\mathbf{x})\varphi_j(\mathbf{x})\,d\mathbf{x} = \begin{cases} 1, & l = j, \\ 0, & l \neq j \end{cases} \qquad (14')$$

une base orthonormée.

D'une manière analogue à (9), on trouve la fonction delta lissée de plusieurs variables

$$\delta_m(\mathbf{x}, \mathbf{z}) = \mathbf{\Phi}'_m(\mathbf{x})\mathbf{\Phi}_m(\mathbf{z}) = \sum_{k=1}^{m}\varphi_k(\mathbf{x})\varphi_k(\mathbf{z}) \qquad (15)$$

et l'on obtient l'estimation de la densité de probabilité multidimensionnelle sous la forme

$$\hat{w}_n(\mathbf{x}, \mathbf{y}_1, \ldots, \mathbf{y}_M) = \frac{1}{M} \sum_{i=1}^{M} \delta_m(\mathbf{x}, \mathbf{y}_i) =$$

$$= \frac{1}{M} \sum_{i=1}^{M} \sum_{k=1}^{m} \varphi_k(\mathbf{x}) \varphi_k(\mathbf{y}_i). \tag{16}$$

**A.1.2. Méthode de Parzen-Nadaraïa [2, 3].** Soit $\mathbf{y} = (y_1, \ldots, y_M)$ un échantillon indépendant d'une distribution inconnue de densité $w_1(x)$. Ecrivons l'estimation $\hat{w}_1(x, \mathbf{y})$ de la densité inconnue comme suit:

$$\hat{w}_1(x, \mathbf{y}) = \frac{1}{Mh(M)} \sum_{i=1}^{M} K\left[\frac{x - y_i}{h(M)}\right], \tag{17}$$

où $K(z)$ est une fonction de pondération quelconque devant satisfaire aux limitations naturelles suivantes:

$$0 \leqslant K(z) < \infty, \quad \int_{-\infty}^{\infty} K(z)\, dz = 1. \tag{18}$$

Si de plus

$$\lim_{|z| \to \infty} zK(z) = 0, \tag{19}$$

$$\lim_{M \to \infty} h(M) = 0, \tag{19'}$$

en un point quelconque de la continuité de $w_1(x)$ l'estimation (17) tend vers $w_1(x)$ en moyenne quadratique.

Par exemple, la fonction

$$K(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \tag{20}$$

satisfait à ces conditions.

L'estimation (17) peut être généralisée au cas multidimensionnel. Soit $\mathbf{y}_1, \ldots, \mathbf{y}_M$ une suite de vecteurs indépendants de dimension $n$ appartenant à la distribution multidimensionnelle $w_n(\mathbf{x})$.

Pour estimer la densité $w_n(\mathbf{x})$ de dimension $n$ on peut utiliser la statistique suivante:

$$\hat{w}_n(\mathbf{x}, \mathbf{y}_1, \ldots, \mathbf{y}_M) =$$

$$= \left[M \prod_{i=1}^{n} h_i(M)\right]^{-1} \sum_{l=1}^{M} \left\{\prod_{i=1}^{n} K\left[\frac{x_i - y_{il}}{h_i(M)}\right]\right\}, \tag{21}$$

où $y_{il}$ est le $i$-ième élément du vecteur échantillonné $\mathbf{y}_l$, $l = 1, \ldots$ $\ldots, M$, appartenant à la distribution $w_n(\mathbf{x})$. Si la fonction $K(z)$ satisfait aux conditions (18), (19), (19′) et si de plus

$$\lim_{M\to\infty} \prod_{i=1}^{n} h_i(M) = 0, \tag{22}$$

$$\lim_{M\to\infty} M \prod_{i=1}^{n} h_i(M) = \infty, \tag{22'}$$

on a également

$$\lim_{M\to\infty} m_1\{|\hat{w}_n(\mathbf{x}, \mathbf{y}_1, \ldots, \mathbf{y}_M) - w_n(\mathbf{x})|^2\} = 0 \tag{23}$$

en chaque point de continuité de $w_n(\mathbf{x})$.

*ANNEXE 2*

## APPROXIMATION STOCHASTIQUE

**A.2.1. Estimation de la racine de l'équation de régression.** Soit $x$ une variable aléatoire répartie suivant la loi $F_1(x \mid \vartheta)$ dépendant du paramètre $\vartheta$ et de variance finie, de plus

$$m_1\{x \mid \vartheta\} = m(\vartheta) < \infty \tag{1}$$

et $m(\vartheta)$ est une fonction monotone croissante de $\vartheta$. On demande d'estimer la racine $\vartheta = \vartheta^*$ *de l'équation de régression*

$$m(\vartheta) = c \tag{2}$$

en supposant que cette racine existe et que les fonctions $F_1(x \mid \vartheta)$ et $m(\vartheta)$ sont inconnues. H. Robbins et S. Monro ont proposé [12] la procédure itérative dite d'*approximation stochastique* permettant d'obtenir une estimation de la racine $\vartheta^*$ à l'aide d'une suite d'observations successives $x_1, \ldots, x_n$, où $x_k$ est la valeur observée de la variable aléatoire $x$ pour $\vartheta = \vartheta_k$, c'est-à-dire

$$m_1\{x_k \mid \vartheta_k\} = m(\vartheta_k), \quad k = 1, \ldots, n. \tag{3}$$

Soient $\vartheta_1$ un nombre constant quelconque et $\{a_n\}$ une suite de nombres positifs quelconques satisfaisant aux conditions suivantes:

$$\sum_{n=1}^{\infty} a_n = \infty, \qquad \sum_{n=1}^{\infty} a_n^2 < \infty. \tag{4}$$

(Par exemple, la suite de termes de la série harmonique $a_n = 1/n$ satisfait aux conditions (4).) Le premier élément $x_1$ de l'échantillon est la valeur observée de $x$ pour $\vartheta = \vartheta_1$. On suppose ensuite $\vartheta_2 = \vartheta_1 + a_1(c - x_1)$. Pour $\vartheta = \vartheta_2$ on observe de nouveau la variable

aléatoire $x$ et l'on obtient l'élément $x_2$ de l'échantillon. L'étape suivante consiste à calculer la valeur $\vartheta_3 = \vartheta_2 + a_2 (c - x_2)$.

En continuant la procédure d'itération proposée on obtient la relation de récurrence suivante *):

$$\vartheta_{n+1} = \vartheta_n + a_n (c - x_n). \tag{5}$$

Il est montré dans [12] que lorsque la taille de l'échantillon d'apprentissage augmente indéfiniment, la grandeur aléatoire $\vartheta_{n+1}$ définie par (5) tend en probabilité vers la racine $\vartheta^*$ de l'équation (2), c'est-à-dire pour $\varepsilon > 0$ on a

$$\lim_{n \to \infty} P\{|\vartheta_{n+1} - \vartheta^*| \geqslant \varepsilon\} = 0. \tag{6}$$

Ainsi, pour $n$ donné la grandeur $\vartheta_{n+1}$ définie par (5) peut être considérée comme une estimation consistante de la racine de l'équation de régression. Pour $c = 0$ la relation (5) donne l'estimation de la position sur l'axe $\vartheta$ du zéro de la fonction de régression $m(\vartheta)$.

En vertu de (5), la suite de grandeurs aléatoires $\{\vartheta_{n+1} - \vartheta_n\}$ est indépendante, car la suite de grandeurs aléatoires $\{x_n\}$ est indépendante. Par conséquent, la suite d'estimations $\{\vartheta_n\}$ est une suite aléatoire à accroissements indépendants.

Notons que la première des conditions (4) laisse un libre choix de la valeur initiale $\vartheta_1$ (on peut la choisir aussi éloignée que l'on veut de $\vartheta^*$), tandis que la seconde condition assure la convergence de $\vartheta_n$ vers $\vartheta^*$ pour $n \to \infty$.

Pour certaines limitations supplémentaires imposées à la fonction de régression, l'estimation $\vartheta_{n+1}$ tend en moyenne quadratique (pour $n \to \infty$) vers la racine de l'équation (2) (voir [8]). Dans [1, 3] la procédure de Robbins-Monro est généralisée au cas multidimensionnel. Le cas limite, lorsqu'on résout le problème en présence d'une réalisation d'apprentissage continue, a été envisagé par nombre d'auteurs (voir [4 à 6]). Les méthodes permettant d'accélérer la convergence de l'approximation stochastique ont été envisagées dans [9, 11, 14]. Le cas où la racine cherchée $\vartheta^*$ varie durant la procédure d'itération a été envisagé dans [7]. Un exposé systématique de la théorie de l'approximation stochastique est donné dans [2].

**A.2.2. Estimation de l'extrémum de la fonction de régression.** On peut utiliser une procédure itérative analogue à (5) pour estimer l'extrémum de la fonction unimodale de régression $m(\vartheta)$, c'est-à-dire pour estimer la racine unique $\vartheta^*$ de l'équation

$$dm(\vartheta)/d\vartheta = 0. \tag{7}$$

*) Si $m(\vartheta)$ est une fonction monotone décroissante, on remplacera le signe plus devant $a_n$ par le signe moins.

Cette procédure a été proposée par J. Kiefer et J. Wolfowitz [10]. Soient $\{a_n\}$ et $\{c_n\}$ des suites de nombres positifs satisfaisant aux conditions

$$\sum_{n=1}^{\infty} a_n = \infty, \quad \sum_{n=1}^{\infty} a_n c_n < \infty, \quad \sum_{n=1}^{\infty} \left(\frac{a_n}{c_n}\right)^2 < \infty, \tag{8}$$

et supposons que l'on ait un échantillon d'apprentissage indépendant $x_1, \ldots, x_{2n}$ pour lequel

$$m_1\{x_{2k} \mid \vartheta_k + c_k\} = m(\vartheta_k + c_k), \tag{9}$$

$$m_1\{x_{2k-1} \mid \vartheta_k - c_k\} = m(\vartheta_k - c_k), \quad k = 1, \ldots, n. \tag{9'}$$

Tout comme dans le problème de l'estimation de la racine de l'équation de régression, la valeur initiale $\vartheta_1$ est quelconque. La procédure itérative de l'estimation de la position sur l'axe $\vartheta$ de l'extrémum de la fonction de régression, est donnée par la relation suivante :

$$\vartheta_{n+1} = \vartheta \pm \frac{a_n}{c_n}(x_{2n} - x_{2n-1}), \tag{10}$$

où le signe plus correspond au cas du maximum de la fonction de régression, et le signe moins au minimum. Il est montré dans [10] que lorsque la taille de l'échantillon d'apprentissage augmente indéfiniment, la variable aléatoire $\vartheta_{n+1}$ définie par (10) tend en probabilité vers la racine $\vartheta^*$ de l'équation (7), celle-ci donnant la position de l'extrémum de la fonction de régression $m(\vartheta)$ sur l'axe $\vartheta$. Les conditions assurant la convergence en moyenne quadratique sont données dans [8]. Dans [3], on trouvera une généralisation au cas multidimensionnel discret, et dans [13] au cas multidimensionnel continu.

Les algorithmes d'approximation stochastique envisagés peuvent être également utilisés pour l'estimation du zéro ou de l'extrémum d'une fonction de régression de la forme $m_1\{f(x) \mid \vartheta\}$, où $f(x)$ est une fonction donnée de la variable aléatoire répartie suivant $F_1(x \mid \vartheta)$. Dans ce cas il y a lieu de remplacer dans (5) et (10) les échantillons d'apprentissage $x_k$ par $f(x_k)$.

*ANNEXE 3*

# TABLES

## A.3.1. Tables des valeurs du seuil $|C|$ dans l'algorithme de signe

| $n$ | $\alpha$ | | | $n$ | $\alpha$ | | | $n$ | $\alpha$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0,025 | 0,01 | 0,005 | | 0,025 | 0,01 | 0,005 | | 0,025 | 0,01 | 0,005 |
| 10 | 8 | 9 | 9 | 30 | 20 | 21 | 22 | 50 | 32 | 33 | 34 |
| 11 | 9 | 9 | 10 | 31 | 21 | 22 | 23 | 52 | 33 | 34 | 35 |
| 12 | 9 | 10 | 10 | 32 | 22 | 23 | 23 | 54 | 34 | 35 | 36 |
| 13 | 10 | 11 | 11 | 33 | 22 | 23 | 24 | 56 | 35 | 37 | 38 |
| 14 | 11 | 11 | 12 | 34 | 23 | 24 | 24 | 58 | 36 | 38 | 39 |
| 15 | 11 | 12 | 12 | 35 | 23 | 24 | 25 | 60 | 38 | 39 | 40 |
| 16 | 12 | 13 | 13 | 36 | 24 | 25 | 26 | 62 | 39 | 40 | 41 |
| 17 | 12 | 13 | 14 | 37 | 24 | 26 | 26 | 64 | 40 | 41 | 42 |
| 18 | 13 | 14 | 14 | 38 | 25 | 26 | 27 | 66 | 41 | 42 | 43 |
| 19 | 14 | 14 | 15 | 39 | 26 | 27 | 27 | 68 | 42 | 44 | 45 |
| 20 | 14 | 15 | 16 | 40 | 26 | 27 | 28 | 70 | 43 | 45 | 46 |
| 21 | 15 | 16 | 16 | 41 | 27 | 28 | 29 | 72 | 44 | 46 | 47 |
| 22 | 16 | 16 | 17 | 42 | 27 | 28 | 29 | 74 | 45 | 47 | 48 |
| 23 | 16 | 17 | 18 | 43 | 28 | 29 | 30 | 76 | 47 | 48 | 49 |
| 24 | 17 | 18 | 18 | 44 | 28 | 30 | 30 | 78 | 48 | 49 | 50 |
| 25 | 17 | 18 | 19 | 45 | 29 | 30 | 31 | 80 | 49 | 50 | 51 |
| 26 | 18 | 19 | 19 | 46 | 30 | 31 | 32 | 85 | 52 | 53 | 54 |
| 27 | 19 | 20 | 20 | 47 | 30 | 31 | 32 | 90 | 54 | 56 | 57 |
| 28 | 19 | 20 | 21 | 48 | 31 | 32 | 33 | 95 | 57 | 59 | 60 |
| 29 | 20 | 21 | 21 | 49 | 31 | 33 | 33 | 100 | 60 | 62 | 63 |

L'hypothèse $H$ est rejetée si le nombre de signes + est supérieur à celui de la table (voir 2.3.1).

La table donne également deux limites $|C|$ et $n - |C|$ dans l'algorithme bilatéral de signe pour la valeur double (voir 2.3.2).

## A.3.2. Distribution de la statistique signe-rang

Pour des valeurs données de $n$ et $x$ la table indique les probabilités (voir 2.4.1)

$$\alpha = P\{S_n(x) \leqslant x\} = P\left\{ S_n(x) > \frac{n(n+1)}{2} - x \right\}$$

| $n$ | $x$ | $\alpha$ | $x$ | $\alpha$ | $x$ | $\alpha$ | $x$ | $\alpha$ |
|---|---|---|---|---|---|---|---|---|
| 10 | 10 | 0,0420 | 8 | 0,0244 | 5 | 0,0098 | 3 | 0,0049 |
| 11 | 13 | 0,0415 | 10 | 0,0210 | 7 | 0,0093 | 5 | 0,0049 |
| 12 | 17 | 0,0461 | 13 | 0,0212 | 9 | 0,0081 | 7 | 0,0046 |

*Suite*

| $n$ | $x$ | $\alpha$ | $x$ | $\alpha$ | $x$ | $\alpha$ | $x$ | $\alpha$ |
|---|---|---|---|---|---|---|---|---|
| 13 | 21 | 0,0471 | 17 | 0,0239 | 12 | 0,0085 | 9 | 0,0040 |
| 14 | 25 | 0,0453 | 21 | 0,0247 | 15 | 0,0083 | 12 | 0,0043 |
| 15 | 30 | 0,0473 | 25 | 0,0240 | 19 | 0,0090 | 15 | 0,0042 |
| 16 | 35 | 0,0467 | 29 | 0,0222 | 23 | 0,0091 | 19 | 0,0046 |
| 17 | 41 | 0,0492 | 34 | 0,0224 | 27 | 0,0087 | 23 | 0,0047 |
| 18 | 47 | 0,0494 | 40 | 0,0242 | 32 | 0,0091 | 27 | 0,0045 |
| 19 | 53 | 0,0478 | 46 | 0,0247 | 37 | 0,0090 | 32 | 0,0047 |
| 20 | 60 | 0,0487 | 52 | 0,0242 | 43 | 0,0096 | 37 | 0,0047 |
| 22 | 75 | 0,0492 | 65 | 0,0231 | 55 | 0,0095 | 48 | 0,0046 |
| 24 | 91 | 0,0475 | 81 | 0,0245 | 69 | 0,0097 | 61 | 0,0048 |
| 26 | 110 | 0,0497 | 98 | 0,0247 | 84 | 0,0095 | 75 | 0,0047 |
| 28 | 130 | 0,0496 | 116 | 0,0239 | 101 | 0,0096 | 91 | 0,0048 |
| 30 | 151 | 0,0481 | 137 | 0,0249 | 120 | 0,0098 | 109 | 0,0050 |
| 32 | 175 | 0,0492 | 159 | 0,0249 | 140 | 0,0097 | 128 | 0,0050 |
| 34 | 200 | 0,0488 | 182 | 0,0242 | 162 | 0,0098 | 148 | 0,0048 |
| 36 | 227 | 0,0489 | 208 | 0,0248 | 185 | 0,0096 | 171 | 0,0050 |
| 38 | 256 | 0,0493 | 235 | 0,0247 | 211 | 0,0099 | 194 | 0,0048 |
| 40 | 286 | 0,0486 | 264 | 0,0249 | 238 | 0,0100 | 220 | 0,0049 |
| 42 | 319 | 0,0496 | 294 | 0,0245 | 266 | 0,0098 | 247 | 0,0049 |
| 44 | 353 | 0,0495 | 327 | 0,0250 | 296 | 0,0097 | 276 | 0,0049 |
| 46 | 389 | 0,0497 | 361 | 0,0249 | 328 | 0,0098 | 307 | 0,0050 |
| 48 | 426 | 0,0490 | 396 | 0,0244 | 362 | 0,0099 | 339 | 0,0050 |
| 50 | 466 | 0,0495 | 434 | 0,0247 | 397 | 0,0098 | 373 | 0,0050 |

### A.3.3. Distribution de la statistique de Kendall

Pour des valeurs données de $n$ et $\alpha$ la table indique les valeurs minimales de $C$ satisfaisant à l'inégalité (voir 2.5.5)

$$P\{S_n(x, y) \geqslant C\} \leqslant \alpha$$

| $n$ | $\alpha$ | | | | | $n$ | $\alpha$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0,005 | 0,010 | 0,025 | 0,050 | 0,100 | | 0,005 | 0,010 | 0,025 | 0,050 | 0,100 |
| 10 | 29 | 27 | 23 | 21 | 17 | 22 | 91 | 83 | 71 | 61 | 47 |
| 11 | 33 | 31 | 27 | 23 | 19 | 24 | 104 | 94 | 80 | 68 | 54 |
| 12 | 38 | 36 | 30 | 26 | 20 | 26 | 117 | 107 | 91 | 77 | 61 |
| 13 | 44 | 40 | 34 | 28 | 24 | 28 | 130 | 118 | 100 | 86 | 68 |
| 14 | 47 | 43 | 37 | 33 | 25 | 30 | 145 | 131 | 111 | 95 | 75 |
| 15 | 53 | 49 | 41 | 35 | 29 | 32 | 160 | 144 | 122 | 104 | 82 |
| 16 | 58 | 52 | 46 | 38 | 30 | 34 | 175 | 157 | 133 | 113 | 89 |
| 7 | 64 | 58 | 50 | 42 | 34 | 36 | 190 | 172 | 146 | 122 | 96 |
| 18 | 69 | 63 | 53 | 45 | 37 | 38 | 205 | 185 | 157 | 133 | 105 |
| 19 | 75 | 67 | 57 | 49 | 39 | 40 | 222 | 200 | 170 | 144 | 112 |
| 20 | 80 | 72 | 62 | 52 | 42 | | | | | | |

# BIBLIOGRAPHIE

## INTRODUCTION


1. R. BELLMAN, R. KALABA — Dynamic programming and modern control theory. Academic Press, N. Y.— London.
2. E. S. VENTZEL — Recherche opérationnelle (en russe). Ed. « Sovetskoe radio », Moscou, 1972.
3. P. M. WOODWARD — Probability and information theory, with application to radar. Pergamon press, 1954.
4. Y. B. GUERMEYER — Introduction à la recherche opérationnelle (en russe). Ed. « Naouka », Moscou, 1971.
5. M. H. De GROOT — Optimal statistical decisions. Mc Graw-Hill, N. Y., 1970.
6. A. N. KOLMOGOROV — Notions fondamentales de la théorie des probabilités (en russe). 2e éd. « Naouka », Moscou, 1974.
7. A. N. KOLMOGOROV — Théorie des probabilités (en russe). Recueil « Mathematica, eïe soderjanié, methody i znatchénié », tome 2, éd. de l'Académie des Sciences de l'U.R.S.S., Moscou, 1956.
8. V. A. KOTELNIKOV — Théorie potentielle de l'insensibilité aux bruits (en russe). Ed. « Gosénergoïzdat », Moscou, 1956.
9. N. N. MOÏSÊEV — Méthodes numériques de la théorie des systèmes optimaux (en russe). Ed. « Naouka », Moscou, 1971.
10. L. S. PONTRIAGUINE, V. G. BOLTIANSKI, R. V. GAMKRÊLIDZÊ, E. F. MICHTCHENKO — Théorie mathématique des processus optimaux Ed. « Mir », Moscou, 1974.
11. A. A. FELDBAUM — Principes théoriques des systèmes asservis optimaux. Ed. « Mir », Moscou, 1973.
12. Ia. Z. TSIPKINE — Adaptation et apprentissage des systèmes automatiques (en russe). Ed. « Naouka », Moscou, 1968.
13. R. CARNAP — Logical foundation of probability. Univ. of Chicago, 1962.
14. U. GRENANDER — Computational probability and statistics. « SIAM Rev.», 1973, n° 1.
15. L. J. SAVAGE — The foundations of statistics. John Wiley, N. Y., 1954.
16. L. A. ZADEH — Outline of a new approach to the analysis of complex systems and decision processes. IEEE Trans., SMCI, 3, 1973, Jan.


## CHAPITRE PREMIER

Le second tome du présent ouvrage est consacré à la théorie statistique des systèmes informationnels pour une information a priori complète. Dans [1] on trouvera un aperçu des résultats essentiels de cette théorie. Les critères de qualité des algorithmes dans les conditions d'indétermination paramétrique (algorithmes uniformément les plus puissants, absence de biais, invariance, similitude, minimax) sont étudiés en détail dans l'ouvrage de E. Lehmann

[5]. L'ouvrage de D. Fraser [12] est consacré aux méthodes de la statistique non paramétrique. Le coefficient d'efficacité asymptotique relative a été introduit par Pitmann dans [4].

Le problème du choix asymptotiquement optimal d'une décision a été pour la première fois envisagé par A. Wald dans [15]. Le développement de la théorie des règles asymptotiquement optimales a été notablement influencé par les travaux de L. Le Cam [13, 14] consacrés à l'étude du comportement asymptotique d'une suite de distribution lors du rapprochement de l'hypothèse et de l'alternative; l'auteur a également donné les conditions d'équivalence asymptotique de ces suites. Il y a lieu de noter également l'article de D. Tchibissov [9].

1. I. A. BOLCHAKOV, L. S. GOUTKINE, B. R. LÉVINE, R. L. STRATONOVITCH — Méthodes mathématiques de la radio-électronique moderne (en russe). Ed. « Sovetskoe radio », 1968.
2. G. P. TARTAKOVSKI (réd.) — Questions de la théorie statistique du radar (en russe), tome 2. Ed. « Sovetskoe radio », 1963.
3. Th. KAILATH — The innovation approach to detection and estimation theory. Proceedings of the IEEE, vol. 58, n° 5, May 1970.
4. M. G. KENDALL, A. STUART — Inference and relationship. The advanced theory of statistics, vol. 2. Griffin, London, 1967.
5. E. L. LEHMANN — Testing statistical hypothesis. John Wiley, N. Y., 1959.
6. D. MIDDLETON — Topics in communication theory. Mc Graw Hill, New York, 1960.
7. D. MIDDLETON, R. ESPOSITO — New results in the theory of simultaneously optimum detection and estimation of signals in noise. Tech. Memo., T-838, Raythe Co., Oct. 1969.
8. R. L. STRATONOVITCH, Yu. G. SOSSOULINE — Détection optimale d'un processus de Markov noyé dans le bruit (en russe). « Izvestia Akad. Naouk SSSR », 1964, n°6.
9. D. M. TCHIBISSOV — Théorème des critères admissibles et son application à un problème de la vérification asymptotique des hypothèses (en russe). « Théoria véroïatnostei i eïe priménénia », 1967, 12, n° 1.
10. C. W. HELSTROM — Statistical theory of signal detection. Pergamon press, 1960.
11. R. E. KALMAN, R. BUCY — New results in linear filtering and prediction theory. Trans. ASME, 1963, March.
12. D. A. S. FRASER — Nonparametric methods in statistics. John Wiley, New York, 1957.
13. L. LE CAM — On the asymptotic theory of estimation and testing hypotheses. Proc. 3. Berkeley Simp. Math. Stat., 1956.
14. L. LE CAM — Locally asymptotically normal families of distributions. Univ. Calif. Publ. Statist., 1960, Pt. 3, 37-98.
15. A. WALD — Asymptotically most powerfull tests of statistical hypotheses. « Ann. Math. Statist.», 1941, v. 12, n°1.

## CHAPITRE 2

La théorie de la vérification des hypothèses sur la forme de fonctions de répartition doit beaucoup aux travaux fondamentaux de A. Kolmogorov et N. Smirnov [29, 14, 15]. Pour une première étude des méthodes de la statistique non paramétrique on peut recommander l'ouvrage d'introduction de G. Noether [33]. Pour une étude plus détaillée de ces méthodes voir D. Fraser [26], J. Bredley [20], M. Puri et P. Sen [34]. Dans l'ouvrage de J. Hajek et Z. Sidak [5] on trouve un exposé complet et systématique de la théorie des statistiques de rang, ainsi qu'un examen détaillé

de la littérature. Les applications pratiques des algorithmes non paramétriques de vérification des hypothèses statistiques aux problèmes de la détection des signaux noyés dans des bruits sont traitées dans l'ouvrage [17] (chap. 8) et l'article de D. Thomas [18].

Les statistiques de rang mixtes ont été proposées et étudiées par E. Feustal et L. Davisson dans [24, 25]. Le détecteur signe-quantile et le corrélateur de coïncidence des polarités modifié sont étudiés dans les ouvrages de D. Laïniotis [30, 31]. Un algorithme de rang simple a été proposé par D. Lomakine [13]. L'ouvrage [36] est consacré à l'étude de l'efficacité du corrélateur de coïncidence des polarités pour la détection d'un signal stochastique. et l'ouvrage [27] à l'efficacité d'un détecteur utilisant la statistique de Kendall. Les algorithmes non paramétriques séquentiels ont été étudiés dans [19]. L'influence de la corrélation des bruits sur les algorithmes de signe et de rang a été envisagée dans [23, 37, 38].

1. T. W. ANDERSON — An introduction to multivariate statistical analysis. John Wiley, New York, 1958.
2. L. N. BOLCHEV — Convertisseurs asymptotiques de Pearson (en russe). « Teoria veroïatnosteï i eïe primenenia », 1963, fasc. 2.
3. L. N. BOLCHEV, N. V. SMIRNOV — Tables de statistique mathématique (en russe). Centre de calcul de l'Académie des sciences de l'U.R.S.S., 1968.
4. B. L. VAN DER WAERDEN — Mathematische statistik. Berlin, 1957.
5. J. HAJEK, Z. SIDAK — Theory of rank tests. Academia Publishing house of the Czechoslovak Academy of sciences. Prague, 1967.
6. B. V. GNÉDENKO — The theory of probability. Chelsea Publishing Co, New York, 1962.
7. D. DUGUÉ — Traité de statistique théorique et appliquée. Masson, 1958.
8. I. A. IBRAGUIMOV, Y. V. LINNIK — Variables indépendantes et stationnairement liées (en russe). Ed. « Naouka », Moscou, 1965.
9. A. M. KAGAN, Y. V. LINNIK, S. R. RAO — Problèmes caractéristiques de la statistique mathématique (en russe). Ed. « Naouka», 1972.
10. M. G. KENDALL, A. STUART — Inference and relationship. The advanced theory of statistics, vol 2. Griffin, London, 1967.
11. H. CRAMER — Mathematical methods of statistics. Princeton Univ. press, Princeton, 1946.
12. E. L. LEHMANN — Testing statistical hypotheses. John Wiley, New York, 1959.
13. D. V. LOMAKINE — Efficacité de certains algorithmes de rang facilement réalisables de détection des signaux noyés dans des bruits (en russe). « Radiotechnica i electronica », 1972, n° 9.
14. N. V. SMIRNOV — Estimation de l'écart entre les courbes de distribution empiriques dans deux échantillons indépendants (en russe). « Bulletin de l'Université de Moscou », 1939, n° 2.
15. N. V. SMIRNOV — Lois approchées de distribution des variables aléatoires à partir des données empiriques (en russe). « Uspekhi matematitcheskikh naouk », 1944, n° 10.
16. F. P. TARASSENKO, V. P. CHOULENINE — Sur certaines propriétés d'une procédure non paramétrique de détection des signaux noyés dans des bruits (en russe) «Radiotechnica», 1971, n° 4.
17. Communication theory, ed. by A. V. BALAKRISHNAN, Mc Graw-Hill, N. Y., 1968.
18. J. B. THOMAS — Nonparametric detection. Proceedings of the IEEE, vol. 58, nb. 5, 1970.
19. C. E. ANTONIAK, G. M. DILLARD — A distribution-free sequential probability-ratio test for multiple-resolution element radars. « IEEE trans.», 1968, IT-14, Novem.
20. J. V. BREDLEY — Distribution-free statistical tests. Prentice Hall, New Jersey, 1968.

21. H. CHERNOFF, I. R. SAVAGE — Asymptotic normality and efficiency of certain nonparametric test statistics. « Ann. Math. Statist.», 1958, v. 29.
22. R. F. DALY, C. K. RUSHFORTH — Nonparametric detection of signal of known form in additiv noise. « IEEE Trans.», 1965, IT-11, n° 1.
23. L. D. DAVISSON, E. A. FEUSTEL, J. W. MODESTINO — The effects of dependence on a nonparametric detection. « IEEE Trans.», IT-16, 1970, n° 1.
24. E. A. FEUSTAL, L. D. DAVISSON — The asymptotic relative efficiency of mixed statistical tests. « IEEE Trans.», IT-13, 1967, n° 2.
25. E. A. FEUSTAL, L. D. DAVISSON — On efficiency of mixed most locally powerful rank tests. « IEEE Trans.», IT-14, 1968, n° 5.
26. D. A. S. FRASER — Nonparametric statistic. J. Wiley, N. Y., 1957.
27. R. A. GROENEVELD — A nonparametric rank correlation method for detection signals in additiv noise. « IEEE Trans.», IT-13, 1967, n° 2.
28. M. KAC, J. KIEFER, J. WOLFOWITZ — On tests of normality and other tests of goodness of fit based on distance methods. « Ann. Math. Statist.», 1955, v. 26, 189.
29. A. KOLMOGOROFF — Sulla determinazione empirica di una logge di distribuzione. Giorn. Inst. Ital. Attuari, 1933, Pt. 4.
30. D. G. LAÏNIOTIS — Generalized distribution-free coincidence detection procedure. « IEEE Intern. Conv. Rec.», 1965, Pt. 7.
31. D. G. LAÏNIOTIS — On a class of generalised PCC. « IEEE Trans.», IT-14, 1968, n° 4.
32. E. LEHMANN — Statistical methods based on ranks. Mc Graw Hill, N. Y., 1974.
33. G. E. NOETHER — Nonparametric statistic. J. Wiley, N. Y., 1967.
34. M. PURI, P. K. SEN — Nonparametric methods in multivariate analysis. J. Wiley, N. Y., 1971.
35. A. RENYI — Warscheinlichkeitsrechnung. VEB, Berlin, 1973.
36. S. S. WOLF, J. B. THOMAS, T. R. WILLIAMS — The polarity-coincidence correlator. « IEEE Trans.», IT-8, 1962, n° 1.
37. S. S. WOLF, J. L. GASTWIRTH, H. RUBIN — The effect of autoregressive dependence on a nonparametric test. « IEEE Trans.», IT-13, 1967, n° 2.
38. M. N. WOINSKY, L. KURZ — Nonparametric detection using dependent samples. « IEEE Trans.», IT-16, 1970, n° 3.

## CHAPITRE 3

La théorie des algorithmes asymptotiquement optimaux de vérification des hypothèses statistiques est basée sur les travaux fondamentaux de L. Le Cam [27, 28]. Ces travaux ont servi de point de départ pour J. Hajek [3, 26] et D. Tchibissov [23] qui ont obtenu des résultats importants pour la théorie asymptotique de la vérification des hypothèses statistiques dans le cas où l'échantillon observé est homogène et indépendant. C'est le cas de la détection d'un signal constant noyé dans des bruits indépendants (détection d'un « biais »).

L'auteur du présent ouvrage et ses collaborateurs A. Kouchnir, A. Pinski, A. Rybine et V. Baronkine [5, 6, 8 à 18, 29 à 31] ont donné des généralisations au cas des échantillons non homogènes (non stationnaires) et non corrélés, ce qui correspond aux signaux radio-électriques brouillés de la pratique courante, ainsi qu'au cas du traitement postdétecteur et des observations quantifiées, ces applications ont une grande portée pratique. Certains résultats donnés dans nos travaux ont été obtenus indépendamment par G. Roussas et R. Jonson. Un exposé rigoureux et complet des fondements mathématiques de la théorie envisagée, avec des généralisations ulté-

rieures, est donné dans la monographie de G. Roussas [34] destinée à un lecteur averti. La monographie de V. Likharev [20] est consacrée aux algorithmes numériques de détection des signaux noyés dans des bruits.


1. V. S. ARKHIPOV, B. R. LÊVINE — Sur les algorithmes de rang asymptotiquement optimaux de détection des signaux dans des bruits corrélés (en russe). « Radiotechnica i electronica », 1975, n° 2.
2. V. M. BARONKINE — Groupement asymptotiquement optimal (en russe). « Radiotechnica i electronica », 1972, n° 9.
3. J. HAJEK, Z. SIDAK — Theory of rank tests. Prague, 1967.
4. I. A. IBRAHIMOV, Yu. V. LINNIK — Variables aléatoires et stationnairement liées (en russe). Ed. « Naouka », 1965.
5. A. F. KOUCHNIR, B. R. LÊVINE — Algorithmes de rang optimaux de détection des signaux noyés dans des bruits (en russe). « Problemy peredatchi informatsii », 1968, n° 3.
6. A. F. KOUCHNIR, A. I. PINSKI — Critères asymptotiquement optimaux de vérification des hypothèses pour un échantillon d'observations lié (en russe). « Teoria veroïatnosteï i eïe primenenia », 1971, n° 2.
7. Yu. KOUTOÏANTS — Sur une classe d'algorithmes analogiques asymptotiquement optimaux de discrimination des signaux (en russe). « Radiotechnica i electronica », 1975, n° 4.
8. B. R. LÊVINE, A. F. KOUCHNIR — Algorithmes asymptotiquement optimaux de détection et de discrimination des signaux noyés dans des bruits (en russe). « Radiotechnica i electronica », 1969, n° 2.
9. B. R. LÊVINE, A. F. KOUCHNIR — Algorithmes de rang asymptotiquement optimaux de détection des signaux noyés dans des bruits (en russe). « Radiotechnica i electronica », 1969, n° 2.
10. B. R. LÊVINE, A. K. RYBINE — Méthodes non paramétriques d'amplitude et de phase de détection des signaux (en russe). I — « Izvestia Acad. Naouk SSSR », sér. cybernétique technique, 1969, n° 5.
11. B. R. LÊVINE, A. K. RYBINE — Méthodes non paramétriques d'amplitude et de phase de détection des signaux (en russe). II — « Izvestia Acad. Naouk SSSR », sér. cybernétique technique, 1970, n° 1.
12. B. R. LÊVINE — Algorithmes optimaux de détection des signaux, insensibles aux variations des données a priori (en russe). « Izvestia vouzov SSSR », sér. radio-électronique, 1970, n° 2.
13. B. R. LÊVINE, A. F. KOUCHNIR — Algorithmes de rang de détection des signaux d'après un échantillon non corrélé (en russe). « Radiotechnica », 1971, n° 4.
14. B. R. LÊVINE, A. F. KOUCHNIR, A. I. PINSKI — Algorithmes asymptotiquement optimaux de détection et de discrimination des signaux noyés dans des bruits corrélés (en russe). « Radiotechnica i electronica », 1971, n° 5.
15. B. R. LÊVINE, V. M. BARONKINE — Détection des signaux noyés dans des bruits d'après des observations quantifiées (en russe). « Radiotechnica i electronica », 1973, n° 5.
16. B. R. LÊVINE — Algorithmes asymptotiquement optimaux de détection des signaux noyés dans des bruits (aperçu, en russe). Travaux du SFTI, 1973, fasc. 63.
17. B. R. LÊVINE, V. M. BARONKINE — Synthèse des algorithmes asymptotiquement optimaux de détection des signaux noyés dans des bruits (en russe). « Radiotechnica i electronica », 1974, n° 5.
18. B. R. LÊVINE — Méthodes asymptotiques dans la théorie statistique des télécommunications. V^e symposium international « Radio-électronique-74 ». Varna, 1974.
19. R. Ch. LIPTZER, A. N. CHIRIAEV — Statistique des processus aléatoires (en russe). Ed. « Naouka », 1974.

20. V. A. LIKHAREV — Méthodes et dispositifs numériques du radar (en russe). Ed. « Sovetskoe radio », 1973.
21. A. K. RYBINE — Classification des signaux faibles noyés dans des bruits (en russe). « Izvestia Acad. Naouk SSSR », sér. cybernétique technique, 1972, n° 5.
22. S. WILKS — Mathematical statistics. John Wiley, N. Y., 1964.
23. D. M. TCHIBISSOV — Théorème des critères admissibles et son application à un problème asymptotique de vérification des hypothèses (en russe). « Teoria veroïatnosteï i eïe primenenia », 1967, n° 1.
24. R. A. JONSON, G. G. ROUSSAS — Asymptotically most powerful test in Marcov processes. « Ann. Math. Statist.», 1964, v. 40, p. 1207.
25. R. A. JONSON, G. G. ROUSSAS — Asymptotically optimal tests in Marcov processes. « Ann. Math. Statist.», 1970, v. 41, p. 918-939.
26. J. HAJEK — Asymptotically most powerful rank-order tests. « Ann. Math. Statist.», 1962, v. 33, p. 1134-1147.
27. L. LE CAM — On the asymptotic theory of estimation and testing hypotheses. Proc. Third Berkeley Symp. Math. Stat., 1956.
28. L. LE CAM — Locally asymptotically normal families of distributions Univ. Calif. Publ. Statist., 1960, Pt. 3.
29. B. R. LEVIN — Einige Aspekte der statistischen Nechrichten-übertragungstheorie. « Nachrichtentechnik ». 1970, n° 5.
30. B. R. LEVIN — Asymptotically optimal technics for testing the nonstationarity of stochastic processes. European meating of statisticians, Budapest, 1972.
31. B. R. LEVIN — Asymptotische Methoden in der statistischen Nachrichtentechnik. « Nachrichtentechnik und Elektronik », 1974, n° 3.
32. G. G. ROUSSAS — Asymptotic inference in Marcov processes. « Ann. Math. Statist.». 1965, v. 36, 978-992.
33. G. G. ROUSSAS — Nonparametric estimation in Marcov processes. Univ. Wisconsin, Madison, 1967.
34. G. G. ROUSSAS — Contiguity of probability measures: some applications in statistics. Cambridge Univ. Press, 1972.
35. P. K. SEN — On the Bachadur representation of sample for sequences of mixing random variables. Journ. Multivariate Analysis, 1972, n° 2.

## CHAPITRE 4

L'exposé systématique des questions liées à la classification des ensembles normaux multidimensionnels est donné dans [2] (voir également [39], [40]). L'algorithme de classification des ensembles normaux avec auto-apprentissage est exposé dans [42, 64]. Les articles [10, 15, 16] sont consacrés à l'étude des propriétés asymptotiques des statistiques de classification lors de l'augmentation du nombre d'indices informationnels. Dans les ouvrages [13, 60] on peut trouver un examen détaillé de la théorie de la reconnaissance des images et de ses applications. On trouvera un aperçu général des derniers ouvrages américains dans [45, 66].

L'exposé des §§ 4.3.2 et 4.3.3 est basé sur [6, 38, 54]. Dans les articles [39, 61] on peut trouver un exposé des procédures semi-euristiques de choix de décisions. Un exposé plus détaillé de la procédure adaptative de Bayes proposée par H. Robbins est donné dans [19, 24]. Un grand nombre d'ouvrages est consacré aux méthodes itératives de construction des algorithmes adaptatifs. Bornons-nous à citer les ouvrages parus ces dernières années [1, 3, 8, 20, 29, 32, 33, 65]; ils donnent ensemble un tableau assez complet de l'état de cette branche de la théorie.

L'ouvrage plus ancien [4] peut servir d'introduction à la théorie de l'adaptation. Dans [8] on trouvera une étude détaillée de la méthode de minimisation du risque empirique, que l'on a omise ici (voir également [49]). La méthode en-

visagée dans 4.3.10 a beaucoup en commun avec celle exposée dans [63]. Pour les algorithmes d'apprentissage sans maître voir également [11, 46, 58, 44]. Les systèmes d'apprentissage à mémoire finie sont étudiés dans [47, 48].

Pour surmonter l'indétermination a priori des paramètres perturbateurs V. Répine et G. Tartakovski [5] ont proposé une méthode adaptative. Ils ont utilisé cette méthode pour l'élaboration des procédures à pas multiples de prise d'une décision [30].

Des algorithmes adaptatifs asymptotiquement optimaux de détection des signaux sont étudiés dans la thèse de A. Pinski [21]. La méthode proposée par Y. Sossouline pour la synthèse des algorithmes adaptatifs asymptotiquement optimaux de détection des signaux noyés dans des bruits, dans le cas d'une indétermination paramétrique et non paramétrique des signaux et des bruits [26 à 28] utilise la corrélation des estimations.

Les propriétés asymptotiques des estimations de Bayes des paramètres du signal sont exposées dans [17, 18, 55].

1. M. A. AÏZERMAN, E. M. BRAVERMAN, L. I. ROSONOER — Méthode des fonctions potentielles dans la théorie de l'apprentissage des machines (en russe). Ed. « Naouka », 1970.
2. T. W. ANDERSON — An introduction to multivariate statistical analysis. John Wiley, New York, 1958.
3. M. AOKI — Optimisation of stochastic systems. Acad. press, 1964, nº 4.
4. R. BELLMAN — Adaptive control processes: a guided tour. Princeton Univ. Press, 1961.
5. I. A. BOLCHAKOV, B. R. LÉVINE, V. G. RÉPINE, G. P. TARTAKOVSKI — Quelques questions de synthèse statistique des systèmes informationnels (en russe). « Izvestia Acad. Naouk SSSR », sér. cybernétique technique, 1970, nº 2.
6. D. BRAVERMAN — Advances in communication systems. Theory and applications. Ed. by A. V. BALAKRISHNAN, Acad. press, 1965, nº 4.
7. M. T. WASAN — Stochastic Approximation. Cambridge Univ. press, 1969.
8. V. N. VAPNIK, A. J. TCHERVONENKIS — Théorie de la reconnaissance des images (en russe). Ed. « Naouka », 1974.
9. O. GOULIACHE — Sur la généralisation de l'algorithme d'apprentissage des fonctions potentielles et la vitesse de convergence (en russe). Ed. « Naouka », 1974.
10. A. D. DÉEV — Représentation des statistiques de l'analyse discriminante et du développement asymptotique dans le cas où la dimension de l'espace est comparable à la taille des échantillons (en russe). « Doklady Acad. Naouk SSSR », 1970, v. 195.
11. A. A. DOROFEÏOUK — Algorithmes d'apprentissage sans maître, basés sur la méthode des fonctions potentielles, proposés pour apprendre aux machines à reconnaître les images (en russe). « Avtomatica i telemecanica », 1966, nº 10.
12. E. B. DYNKINE — Statistiques nécessaires et suffisantes pour une famille de répartitions des probabilités (en russe). « Uspekhi mathematitcheskikh naouk, 1951, nº 1.
13. N. G. ZAGOROUÏKO — Méthodes de reconnaissance et applications (en russe) Ed. « Sovetskoe radio », 1972.
14. H. CRAMER — Mathematical methods of statistics. Princeton Univ. press, Princeton, 1946.
15. B. R. LÉVINE, E. B. TROÏTSKI — Probabilité totale d'erreur lors de la classification des ensembles normaux, différant par leurs vecteurs des moyennes (en russe). « Avtomatica i telemecanica », 1970, nº 1.
16. B. R. LÉVINE, E. B. TROÏTSKI — Sur l'accumulation d'indices dans les problèmes de classification des observations (en russe). « Radiotechnica i electronica », 1970, nº 7.

17. B. R. LÉVINE, Yu. S. CHINAKOV — Propriétés asymptotiques de l'estimation de Bayes du coefficient de régression en présence d'un bruit non stationnaire (en russe). « Problemi peredatchi informatsii », 1967, n° 3.
18. B. R. LÉVINE, Yu. S. CHINAKOV — Forme limite de l'estimation de Bayes des coefficients de régression en présence d'un bruit non stationnaire (en russe). « Problemi peredatchi informatsii », 1967, n° 1.
19. J. NEYMAN — Two break-throughs in the theory of statistical decision making. « Revue de l'Inst. Intern. de Stat.», 30, n° 1 (1962), PU-27.
20. M. B. NEVELSON, R. Z. KHASMINSKI — Approximation stochastique et estimation récurrente (en russe). Ed. « Naouka », 1972.
21. A. I. PINSKI — Algorithmes asymptotiquement optimaux de détection et de discrimination des signaux faibles noyés dans des bruits corrélés (en russe). Exposé de thèse, 1972.
22. V. F. PISSARENKO — Sur l'estimation des paramètres d'un processus stationnaire gaussien (en russe). « Litovski mathematitcheski sbornik », 1962, n° 2.
23. V. S. POUGATCHEV — Algorithmes optimaux d'apprentissage des systèmes automatiques avec un maître non parfait (en russe). « Dokladi Akademii Naouk SSSR », 1967, t. 172, n° 5.
24. H. ROBBINS — The empirical Bayes approach to statistical decision problems. AMS, 35, n° 1 (1964).
25. G. S. SEBESTYEN — Decision-making processes in pattern recognition. New York-London, Mcmillan, 1962.
26. Yu. G. SOSSOULINE — Réception optimale et non optimale des signaux en cas d'information incomplète (en russe). « Radiotechnica i electronica », 1970, n° 12.
27. Yu. G. SOSSOULINE — Principe d'estimations corrélées de réception des signaux noyés dans des bruits et information a priori (en russe). « Radiotechnica i electronica », 1971, n° 3.
28. Yu. G. SOSSOULINE — Méthodes d'estimations corrélées de détection des signaux faibles noyés dans des bruits et synthèse mathématico-euristique (en russe). « Izvestia Akad. Naouk SSSR », sér. cybernétique technique, 1971, n° 5.
29. R. L. STRATONOVITCH — Principe de la réception adaptative (en russe). Ed. « Sovetskoe radio », 1973.
30. G. P. TARTAKOVSKI, V. G. RÉPINE — Adaptation des processus de prise de décision à étapes multiples basée sur la connaissance des pertes aux étapes antérieures (en russe). « Radiotechnica i électronica », 1972, n° 11.
31. S. S. WILKS — Mathematical statistics. John Wiley, N. Y., 1964.
32. L. I. FILIPPOV — Bases de la théorie de la réception des signaux discrets (en russe). Ed. « Naouka », 1974.
33. K. S. FU — Sequential methods in the pattern recognition and machine learning. Academic press, N. Y., 1968.
34. E. J. HANNAN — Multiple time series. John Wiley, N. Y., 1970.
35. Ya. Z. TSIPKINE — Fondements de la théorie des systèmes d'apprentissage (en russe). Ed. « Naouka », 1970.
36. C. SCIBI — Sur l'apprentissage des machines pour des règles itératives avec mémoire (en russe) « Problemi oupravlenia i teorii informatsii », 1972, n° 1.
37. V. V. CHAKHGUILDIAN, M. S. LOKHVITSKI — Méthodes de la réception adaptative des signaux (en russe). Ed. « Sviaz », 1974.
38. N. ABRAMSON, D. BRAVERMAN, G. SEBASTYEN — Pattern recognition and machine learning. « IEEE trans.», IT-9, 1963, n° 4.
39. T. W. ANDERSON — Classification by multivariate analysis. « Psychometrika », 1951, v. 16.
40. T. W. ANDERSON, R. R. BACHADUR — Classification into two multivariate normal distribution with different covariance matrices. « Ann. Math. Statist.», 1962, n° 2.

41. A. H. BEWKER — A representation of Hotelling's and Anderson's classification statistics in term of simple statistics. Stanford Math. in social sciences, 1961.
42. D. B. COOPER, P. W. COOPER — Nonsupervised adaptativ signal detection and pattern recognition. « Inf. and Contr.», 1964, n° 3.
43. D. B. COOPER — Adaptativ pattern recognition and signal detection using stochastic approximation. « IEEE Trans.», ES-13, 1964, n° 3.
44. D. B. COOPER — When should a learning machine ask for help? « IEEE Trans.», IT-20, 1974, n° 4.
45. T. M. COVER — Recent books on pattern recognition. « IEEE Trans.» IT-19, 1973, n° 6.
46. S. C. FRALICK — Learning to recognize pattern without a teacher. « IEEE Trans.», IT-13, 1967, n° 1.
47. M. E. HELLMAN — The effects of randomization on finite-memory decision schemes. « IEEE Trans.», IT-18, 1972, n° 4.
48. M. E. HELLMAN — Finite memory algorithms for estimation the means of a gaussian distribution. « IEEE Trans.», IT-20, 1974, n° 3.
49. W. H. HIGHLEYMAN — Linear decision function with application to pattern recognition. « Proc. IRE », 1962, n° 6.
50. M. J. HINCH — A model to a self-adaptativ filter. « Inform. and Contr.», 1962, Sept.
51. S. JOHN — On some classification statistics. « Indian Journ. of statist.», 1960, v. 22.
52. S. JOHN — Errors in discrimination. « Ann. Math. Statist.», 1961, n° 4.
53. D. G. KABE — Some results on the distribution of two random matrices used in classification procedures. « Ann. Math. Statist.», 1963, n° 2.
54. M. A. KAC — A note on learning signal detection. « IRE Trans.» IT-8, 1962, n° 2.
55. B. R. LEVIN, Yu. S. SHINAKOV — Asymptotic properties of Bayes estimates of parameters of signal masked by interference. « IEEE Trans.», IT-18, 1972, n° 1.
56. T. MARILL, D. GREEN — Effectiveness of receptors in recognition system. « IEEE Trans.», IT-9, 1963, n° 1.
57. M. OKATOMO — Discrimination for variance matrices. « Ann. Math. Statist.», 1961, n° 1.
58. E. A. PATRICK, J. C. HANCOCK — The nonsupervised learning of probability spaces and recognition of patterns. « IEEE Conv. Rec.», 1965.
59. H. J. SCUDDER — Adaptiv communication receivers. « IEEE Trans.» IT-11, 1965, n° 2.
60. J. SEILDER — Optimierung informationsübertagender Systeme. Band 2, VEB Verlag Technik, Berlin, 1969.
61. J. SPRAGINS — Learning without a teacher. « IEEE Trans.» IT-12, 1966, n° 2.
62. G. T. TOUSSAINT — Bibliography on estimation of misclassification. « IEEE Trans.», IT-20, 1974, n° 4.
63. C. T. WOLVERTON, T. J. WAGNER — Asymptotically optimal discriminant functions for pattern classification. « IEEE Trans », IT-15, 1969, n° 2.
64. A. V. MILENKI — Classification des signaux en cas d'indétermination. Ed. « Sovetskoe radio », 1975.
65. A. P. SAGE, J. L. MELSA — System identification. Academic press, N. Y., 1970.
66. L. KANAL — Patterns in pattern recognition: 1968-1974. « IEEE Trans.», IT-20, 1974, n° 6.

## CHAPITRE 5

L'exposé systématique de la théorie des règles semblables et invariantes de vérification des hypothèses composées est donné dans l'ouvrage de E. Lehmann [11]. On peut recommander au lecteur ayant pris connaissance de ce livre de se référer aux monographies [5, 12]. La méthode régulière de recherche de l'invariant maximal a été proposée par V. Kouznetsov [8]. Les principes d'invariance et de similitude sont utilisés dans [1] pour la synthèse des algorithmes de détection d'un signal en présence de paramètres inconnus (perturbateurs).

Le problème de la détection d'un signal noyé dans un bruit de variance inconnue est étudié dans [2, 7, 13, 19]. La thèse de G. Skvortsov [14] est consacrée à l'une des méthodes de solution du problème de la recherche de la règle invariante de détection d'un signal noyé dans un bruit de fonction de corrélation inconnue.

L'algorithme optimal de classification des variables aléatoires unidimensionnelles a été étudié par C. Rao [18]. S. Gupta [16] a donné une généralisation au cas multidimensionnel. P. Dankov [4] a proposé un algorithme de classification invariant uniformément le plus puissant. Dans [3] sont construites les règles de classification semblables asymptotiquement optimales basées sur la théorie exposée au chapitre 3 du présent ouvrage.

1. V. A. BOGDANOVITCH — Méthode de construction des règles semblables de détection des signaux pour une indétermination a priori (en russe). « Radiotechnica i electronica », 1970, n° 7.
2. V. A. BOGDANOVITCH, V. N. PROKOFIEV — Sur la propriété d'invariance du critère non cohérent de détection dans des bruits (en russe). « Radiotechnica i electronica », 1968, n° 2.
3. V. O. GOLTSOV, A. I. PINSKI — Règles de classification semblables asymptotiquement optimales (en russe). « Radiotechnica i electronica », 1975, n° 4.
4. P. P. DANKOV — Procédures optimales d'apprentissage de choix des décisions statistiques (en russe). « Radiotechnica i electronica », 1965, n° 10.
5. A. M. KAGAN, Yu V. LINNIK, S. R. RAO — Problèmes de caractérisation de la statistique mathématique (en russe). Ed. « Naouka », 1972.
6. M. G. KENDALL, A. STUART — Inference and relationship. The advanced theory of statistics, vol 2. Griffin, London, 1967.
7. V. A. KORADO — Sur la détection optimale des signaux noyés dans des bruits de paramètres inconnus pour une probabilité limitée de fausse alarme (en russe). « Radiotechnica i electronica », 1972, n° 1.
8. V. P. KOUZNETSOV — Invariance des décisions vis-à-vis des paramètres perturbateurs (en russe). « Problemy peredatchi informatsii », 1971, n° 4.
9. B. R. LÉVINE, V. P. KOUZNETSOV — Fonctions de décision invariantes et méthode du maximum de vraisemblance pour une situation indéterminée. Communication à la seconde Conférence nationale sur les problèmes statistiques de la théorie de la gestion (en russe). Ed. « Naouka », 1970.
10. B. R. LÉVINE, G. I. SKVORTSOV — Détection d'une composante harmonique noyée dans un bruit normal à paramètres spectraux inconnus (en russe). « Radiotechnica i electronica », 1974, n° 2.
11. E. L. LEHMANN — Testing statistical hypothesis. John Wiley, N. Y., 1959.
12. Yu. V. LINNIK — Problèmes statistiques à paramètres perturbateurs. Ed. « Naouka », 1966.
13. A. L. ROUKHINE, S. V. SAMSONENKO — Sur une procédure de détection invariante par rapport à l'intensité du signal et du bruit (en russe). « Radiotechnica i electronica », 1972, n° 1.

14. G. I. SKVORTSOV — Questions de synthèse statistique d'un système de détection des signaux invariant par rapport au spectre du bruit (en russe). Résumé de doctorat, 1974.
15. Communication theory, ed. by A. V. BALAKRISHNAN. Mc Graw-Hill, N. Y., 1968.
16. S. D. GUPTA — Optimum classification rules for classification into two multivariate normal populations. « Ann. Math. Statist.», 1965, n° 4.
17. M. G. KENDALL, A. STUART — The advanced theory of statistics, vol. 3, Griffin, London, 1968.
18. C. R. RAO — A general theory of discrimination when the information about alternative populations is based on sample. « Ann. Math. Statist.» 1954, n° 4.
19. L. L. SCHARF, D. W. LYTLE — Signal detection in gaussian noise of unknown level. « IEEE Trans. », IT-17, 1971, n° 4.

## ANNEXE 1

1. M. A. AÏZERMAN, E. M. BRAVERMAN, L. I. ROZONOER — Méthode des fonctions potentielles dans la théorie de l'apprentissage des machines (en russe). Ed. « Naouka », 1970.
2. E. J. NADARAÏA — Nouvelles estimations des fonctions de répartition (en russe) « Teoria veroïatnosteï i eïe primenenia », 1957, n° 3.

## ANNEXE 2

1. E. G. GLADYSHEV — Sur l'approximation stochastique (en russe). « Teoria veroïatnosteï i eïe primenenia », 1965, n° 2.
2. M. B. NEVELSON, R. Z. KHASMINSKI — Approximation stochastique et estimation récurrente (en russe). Ed. « Naouka ». 1972.
3. J. A. BLUM — Multidimensional stochastic approximation procedures. « Ann. Math. Statist.», 1965, n° 4.
4. S. CSIBI — On continuous stochastic approximations. Proc. Colloq. Inform. Theory, Budapest, 1967.
5. M. DRIML, N. NEDOMA — Stochastic approximation for continuous random processes. « Trans. Conf. Inform. Theory », Prague, 1960.
6. M. DRIML, O. HANS — Continuous stochastic approximations. « Trans. Conf. Inform. Theory », Prague, 1967.
7. V. DUPAC — A dynamic stochastic approximation method. « Ann. Math. Statist.», 1965, 1665-1702.
8. A. DVOROTZKY — On stochastic approximation method. « Ann. Math. Statist.», 1951, n° 1.
9. V. FABIAN — Stochastic approximation of minima with improved asymptotic speed. « Ann. Math. Statist.», 1967, n° 1.
10. J. KIEFER, J. WOLFOWITZ — Stochastic estimation of the maximum of regression function. « Ann. Math. Statist.», 1952, n° 3.
11. H. KESTEN — Acceleration stochastic approximation. « Ann. Math. Statist.», 1958, n° 1.
12. H. ROBBINS, S. MONRO — A stochastic approximation method. « Ann. Math. Statist.», 1951, n° 1.
13. D. L. SAKRISON — A continuous Kiefer-Wolfowitz procedure for random processes. « Ann. Math. Statist.», 1964, n° 2.
14. J. H. VENTER — An extention of the Robbins-Monro procedure. « Ann. Math. Statist.», 1967, n° 1.

## NOTATIONS UTILISÉES

| | |
|---|---|
| $\mathbf{A}$, $A_{ij}$ | — matrice et ses éléments |
| $a$ | — signal constant, valeur moyenne d'une variable aléatoire |
| $a_s$ | — composante constante d'un signal |
| $B(t, y)$, $B(\tau)$ | — fonction de corrélation |
| $C$ | — seuil |
| $\mathbf{D}$, $D_{ij}$ | — matrice et ses éléments |
| $d_N$ | — « distance » entre les classes |
| $F_n(\mathbf{x})$ | — fonction de répartition |
| $F_1^{-1}(y)$ | — fonction inverse de la fonction de répartition unidi- — mensionnelle |
| $F(x)$ | — intégrale de Laplace |
| $f(x)$ | — caractéristique d'un convertisseur sans inertie |
| $H$ | — hypothèse |
| $h(u, v)$ | — fonction de transition impulsionnelle d'un système linéaire |
| $\mathrm{I}_f$ | — quantité d'information de Fisher |
| $\mathrm{I}_0(x)$ | — fonction de Bessel de l'argument imaginaire |
| $\mathrm{I}_q(x, y)$ | — rapport de la fonction bêta incomplète à la fonction bêta complète |
| $K$ | — alternative |
| $\mathrm{K}$ | — noyau |
| $\mathbf{K}$, $K_{ij}$ | — matrice et ses éléments |
| $\mathbf{k}(z)$ | — fonction de répartition de Kolmogorov |
| $l(\mathbf{x})$ | — rapport de vraisemblance |
| $L_{\mathbf{x}}(\vartheta)$ | — fonction de vraisemblance |
| $m$ | — nombre de fonctions de base, nombre d'éléments d'un groupe dans les statistiques non ordonnées |
| $\mathbf{M}$, $M_{ij}$ | — matrice de corrélation et ses éléments |
| $M_2$ | — moment centré deux |
| $m_1$ | — moment un (moyenne) |
| $n$ | = taille d'un échantillon |
| $N$ | — dimension des classes classifiées |

| | |
|---|---|
| $p$ | — probabilité a priori |
| $P$ | — probabilité d'un événement |
| $q$ | — probabilité a priori |
| $r$ | — coefficient de corrélation |
| $R_i$ | — rang de l'élément $x_i$ d'un échantillon |
| $\mathbf{R}^+$, $\mathbf{R}^-$ | — vecteur des rangs positifs (négatifs) |
| $R$ | — risque moyen |
| $S_1$, $S_2$ | — ensembles classifiés |
| $s(t)$ | — signal |
| $s_i$ | — valeur du signal à l'instant $t_i$ |
| $\mathrm{S}_n(\mathbf{x})$ | — statistique |
| $t$ | — temps courant |
| $t_\alpha$ | — l'$\alpha$-quantile de la loi de Student |
| $t(\mathbf{x}, \mathbf{y})$ | — statistique |
| $T$ | — durée d'une observation |
| $u(x)$ | — fonction du saut unitaire |
| $\mathrm{V}$ | — statistique de classification |
| $\hat{\mathrm{V}}$ | — statistique de classification avec échantillons d'apprentissage |
| $\mathrm{W}_s$ | — puissance d'un signal |
| $w_1(x)$, $w_n(\mathbf{x})$ | — densité de probabilité |
| $\mathbf{x}$ | — vecteur des valeurs échantillonnées |
| $x_i$ | — élément d'un échantillon |
| $x^{(k)}$ | — statistique d'ordre |
| $\mathbf{X}$ | — espace échantillonné |
| $x_\alpha$ | — l'$\alpha$-quantile de la loi normale |
| $x(t)$ | — réalisation d'un processus aléatoire |
| $y_n(\mathbf{x})$ | — statistique |
| $\mathbf{Y}_n$ | — statistique vectorielle |
| $z_n(\mathbf{R})$ | — statistique de rang |
| $\alpha$ | — probabilité de fausse alarme |
| $\beta$ | — probabilité de perte du signal |
| $\Gamma(x)$ | — fonction gamma |
| $\gamma$ | — constante $= \lambda\sqrt{n}$ |
| $\delta(x)$ | — fonction delta |
| $\delta_{ij}$ | — symbole de Kronecker |
| $\delta_n$ | — algorithme de prise d'une décision |
| $\Delta$ | — intervalle de quantification |
| $\lambda$ | — amplitude du signal |
| $\Lambda$ | — rapport de vraisemblance moyenné |
| $\xi$, $\eta$ | — variables aléatoires |
| $\xi(t)$ | — processus aléatoire |
| $\psi(\mathbf{x})$ | — statistique |
| $\mathbf{\Pi}$, $\Pi_{ij}$ | — matrice de pertes et ses éléments |
| $\sigma^2$ | — variance |

$\boldsymbol{\vartheta}$, $\hat{\boldsymbol{\vartheta}}$ — paramètre d'une loi de distribution et son estimation
$\Theta$ — espace des paramètres
$\Theta_1(v)$ — fonction caractéristique d'une variable aléatoire
$\rho$ — coefficient d'efficacité asymptotique relative
$\Phi(\mathbf{x})$ — fonction de décision
$\varphi_k(t)$ — fonction de base d'un signal
$\varphi_{k,i}$ — matrice du signal
$\varphi^c_{k,i}$ — vecteur colonne de la matrice du signal
$\chi^2_\alpha(x, y)$ — l'$\alpha$-quantile de la loi du $\chi^2$ non centrée
$\chi^2_\alpha(x)$ — l'$\alpha$-quantile de la loi du $\chi^2$ centrée
$\chi_k(x_i)$ — indicateur d'un ensemble
sgn $x$ — signe de la grandeur $x$
tr $\mathbf{A}$ — trace de la matrice $\mathbf{A}$
$\binom{n}{k}$ — nombre des combinaisons de $n$ éléments $k$ à $k$
$\in$ — symbole d'appartenance à un ensemble
$\sim$ — égalité asymptotique (pour $n \to \infty$)
$\boldsymbol{\vartheta}'$, $\mathbf{M}'$ — vecteur $\boldsymbol{\vartheta}$ et matrice $\mathbf{M}$ transposés

## Désignations sur les schémas

$\int$ — limiteur
$+$ — addition
$-$ — soustraction
$\times$ — multiplication
$=$ — comparaison
$\underline{\vee}$ — redresseur linéaire
$\Sigma$ — sommateur-accumulateur
$\uparrow^2$ — élévateur au carré
$\sqcup\!\sqcap$ — dispositif à seuil
LR — ligne à retard
Cor — corrélateur
R — dispositif de rangement
$\varphi$ — générateur des fonctions de base
OL — oscillateur local

# INDEX ALPHABÉTIQUE DES MATIÈRES


Algorithme 9
— adaptatif 33, 37, 238
— — de Bayes 231
— — de Robbins 231
— d'apprentissage 21
— asymptotiquement optimal 41, 188
— — le plus efficace 46
— de classification 33, 205, 282, 284
— — avec apprentissage 33
— — avec auto-apprentissage 215
— consistant 20, 37
— de détection 16
— discret 20
— de discrimination 16, 25
— invariant 262, 273, 279
— linéaire 21, 71, 78
— localement optimal 40
— mixte 66
— non paramétrique 32, 66
— non randomisé 21
— optimal de Neyman 260
— postdétecteur 21, 149
— d'amplitude 149
— — de phase 151
— de rang 64, 93, 115
— — à médiane 94
— — simplifié 98
— — de Van der Waerden 95
— — de Wilcoxon 81, 94
— — — à échantillon double 94
— récurrent 21, 234
— semblable 281
— — de Neyman 260
— de signe 63, 67, 71, 74
— — dans les bruits currélés 118
— de signe-rang 66, 80, 84
— — — mixte 87
— — — de Van der Waerden 86
— simple 21
— unilatéral 47, 71
Alternative composite 34

Apprentissage
— avec maître 33, 206, 234
— sans maître 33
Approximation stochastique 294
Auto-apprentissage 215

Bruit 22
— additif 25
— blanc 23
— corrélé 23, 178
— « étranger » 50, 189
— indépendant 65, 144
— laplacien 73
— markovien 181
— — multiplement connexe 188
— multiplicatif 25
— normal 23, 73
— stationnaire 23, 65

Coefficient d'efficacité relative 36, 45, 47
Condition de factorisation 24
Contiguïté 131
Convertisseur 135
Corrélateur 135, 183, 188
— de coïncidence des polarités 108
Critère
— asymptotique 38
— de Bayes 17, 20
— de conformité 52
— — de Kolmogorov 54, 60
— — de Mises 58, 60
— — de Renyi 57
— — de Smirnov, unilatéral 56, 60
— — de Wilcoxon 61
— — du $\chi^2$ 52
— du maximum de la probabilité a posteriori 18, 19
— — de vraisemblance 18, 20, 35
— du minimax 18

Critère du minimum du carré moyen de l'erreur 20
— — du risque moyen 17
— de Neyman-Pearson 18, 34, 38
— de qualité des algorithmes 12, 35, 37, 38
— de stabilité 50

Détecteur asymptotiquement optimal 9
— à deux canaux 100
— linéaire 71
— optimal 28
— de rang 98, 109
— du signal stochastique 100
— de signe, bilatéral 70, 74
— —, unilatéral 67
— signe-quantiles 79
— signe-rang 80, 90
— — — mixte 87
Détection non cohérente 192
Distribution des paramètres du bruit, laplacienne 73, 166
— — —, logistique 166
— — —, normale 73, 135, 165
— — — — limite 51
— — —, sinusoïdale de phase aléatoire 74
— — du signal, normale 115, 192
— — —, de Poisson 232

Echantillon d'apprentissage 9, 33
— — classifié 33, 206
— — non classifié 33, 215
— homogène 65
— indépendant 65, 206
— ordonné 52, 63
— quantifié 171
Efficacité relative 45
— — de l'algorithme de signe 118
— — — signe-rang 84
— — — — — mixte 88
— — — de Wilcoxon 95
— — — — à double échantillon 112
— — du corrélateur de coïncidence des polarités 105
— — du détecteur de signe bilatéral 74
— — — — unilatéral 71
Equation de régression 294
Equivalence asymptotique 132
Estimation 19
— bayesienne 29, 247
— —, propriétés asymptotiques 249
— consistante 38
— de fonctions de vraisemblance 33
Estimation du maximum de vraisemblance 33
— optimale 29

Filtrage 29
— optimal linéaire 30
Fonction d'autocorrélation 192
— bêta 69
— de corrélation 23, 192
— de décision 21
— delta 290
— des pertes 253
— potentielle 291
— de régression 233
— de répartition logistique 166
— — normale 23
— — de Poisson 232
— — de Rayleigh 148
— — — généralisée 148, 150
— — de Wishart 211
— — du $\chi^2$ 158, 191
Fonctionnelle du rapport de vraisemblance 25
Formule de Bayes 223

Groupe des déplacements 35
— des permutations 278
— des rotations 35
— de transformations 34, 262, 274, 278

Hypothèse statistique composite 16, 53
— — non paramétrique 32,
— — simple 16, 53

Indétermination a priori 8, 16
— — non paramétrique 32, 35, 39
— — paramétrique 31, 33
Indices informationnels 217
— — corrélés 219, 221
Information d'après Fisher 129, 135, 154, 172
Intégrale de Laplace 27, 44, 69, 165, 206
Intégrateur parfait 121
Intervalle de quantification 171
Invariance 66
Invariant maximal 262, 268, 274, 279

Limiteur parfait 81, 99
Logarithme du rapport de vraisemblance 25, 128, 178

Logarithme du rapport, développement asymptotique 179

Matrice de corrélation du bruit 27
— — — auto-adjointe 28
— — — hermitienne 28
— diagonale 190
— des pertes 17
Méthode d'approximation stochastique 233
— de classification géométrique 214
— de détection d'amplitude 149
— — de phase 151
— euristique 227
— des fonctions potentielles 236
— de Parzen-Nadaraïa 237

Paramètres perturbateurs 25, 257, 271
Principe d'invariance 34
— du maximum de Pontriaguine 13
— de similitude 34
Probabilité de détection correcte 27, 34, 134, 143
— d'erreur de classification 206, 219, 222
— de fausse alarme 18, 27, 34, 36, 79, 134, 143, 173, 181
— de perte du signal 18, 79, 134, 173
— a priori de présence du signal 17
Problème d'analyse statistique 15
— de détection de signaux 16, 25
— — — et d'estimation des paramètres 17
— de discrimination de signaux 16, 25
Processus aléatoire 15
— — normal 23
— markovien 23, 178
— — de connexité $k$ 24
— — vectoriel 24
— sans post-action
Puissance du bruit 257
— du signal 129

Quantification d'observations en amplitude 20, 171
— optimale 176
— temporelle 20
Quantile de la loi normale 69, 134, 173
— — de Student 74, 269
— — du $\chi^2$ 158, 191

Rang d'un élément d'échantillon 64
Rapport signal/bruit 44, 137, 251
— de vraisemblance 25, 128
Règle admissible 46
— de détection, non biaisée 34, 69
— —, similaire 34
— uniformément la plus puissante 34
Risque moyen 17
— a posteriori 19

Signal 22
— à bande étroite 145
— déterministe 76, 90, 137, 182
— modulé 145, 149, 154
— quasi déterministe 22, 147
— stochastique 23, 158
— vectoriel 22
Statistique asymptotiquement suffisante 41, 133
— centrée 138
— de classification 217, 221
— de Kendall 109
— normée 43
— d'ordre 63
— de rang 64, 94
— de signe 62
— signe-rang 66, 87
— de Spearman 109
— suffisante 26
— vectorielle 186
Suite markovienne 64, 178
Symbol de Kronecker 155, 181
Système informationnel 15

Taille d un échantillon 52
Théorème
— de Bernstein et Mises 230
— de le Cam 132
— de la limite centrale 39, 48, 71, 130
Traitement d'observations analogique 20
— — discret 20
— — numérique 20
— — postdétecteur 21

Vecteur des indices informationnels 217
— de rang 64
— des rangs positifs 66
— de signe 64
— des statistiques d'ordre 63
— transposé 22, 179

# INDEX DES NOMS

Abramson N. 306
Aïzerman M. A. 305, 309
Anderson T. W. 305, 306
Antoniak C. E. 301
Aoki M. 305
Arkhipov V. S. 303

Bachadur R. R. 306
Balakrishnan V. 305
Baronkine V. M. 303
Bayes Th. 8*
Bellman I. A. 30, 300, 305
Berens 111
Bernoulli 105
Bernstein S. N. 230
Bewker A. H. 307
Blum J. A. 309
Bogdanovitch V. A. 308
Bolchakov I. A. 30, 300, 305
Bolchev L. N. 301
Boltianski V. G. 299
Braverman D. 305, 306
Braverman E. M. 305
Bredley V. 301
Bucy R. 30, 300

Carnap R. 299
Chakhguildian V. V. 306
Chernoff H. 302
Chinakov Yu. S. 306
Chiriaev A. N. 303
Choulénine V. P. 301
Cooper D. B. 307
Cooper P. W. 307
Cover T. M. 307
Cramer H. 301, 305
Csibi S. 6, 306, 309

Daly R. F. 302
Dankov P. P. 308
Davisson L. D. 302
De Groot M. H. 299
Déev A. D. 302
Dillard G. H. 301
Dorofeïouk A. A. 305
Driml M. 309
Dugue D. 301
Dupac V. 309
Dvorotzky A. 309

Esposito R. 30, 300

Fabian V. 309
Feldbaum A. A. 299
Feustal E. A. 302
Filippov L. I. 306
Fisher R. A. 111, 151
Fralick S. C. 307
Fraser D. A. S. 300, 302
Fu K. S. 306

Gamkrélidzé R. V. 299
Gastwirth J. L. 302
Gladyshev E. G. 309
Gnédenko B. V. 301
Goltsov V. O. 308
Gouliache O. 305
Goutkine L. S. 300
Green D. 307
Grenander U. 299
Groeneveld R. A. 302

* Ici et dans la suite l'astérisque désigne une mention fréquente du nom.— *Note du rédacteur.*

Guermeyer Y. B. 299
Gupta S. D. 309

Hajek J. 301, 304
Hancock J. 307
Hannan E. J. 306
Hans O. 309
Hellman M. E. 307
Helstrom C. W. 300
Highleyman W. H. 307
Hinch M. J. 303
Hoeffding 66
Hopf 30

Ibrahimov I. A. 301, 303

John S. 307
Jonson R. A. 304

Kabe D. G. 307
Kac M. A. 302, 307
Kagan A. M. 301, 308
Kaïlath Th. 31, 300
Kalman R. E. 30, 300
Kanal L. 307
Kendall M. G. 300, 301, 308, 309
Kesten H. 309
Khasminski R. Z. 306, 309
Kiefer J. 296, 302, 309
Kolmogoroff A. 302
Kolmogorov A. N. 54, 299
Korado B. A. 308
Kotelnikov V. A. 7, 299
Kouchnir A. F. 303
Koutoïants Yu. 303
Kouznetsov V. P. 308
Kronecker L. 155
Kurz L. 302

Laïniotis D. G. 302
Laplace P. S. 44*
Le Cam L. 132, 300, 304
Lehmann E. L. 300-302, 308
Levin B. R. 304, 307
Lévine B. R. 300, 303-306, 308
Likharev V. A. 304
Linnik Yu. V. 301, 303, 308
Liptzer R. Ch. 303
Lokhvitski M. S. 306
Lomakine D. V. 301
Lytle D. W. 309

Mann 61
Marill T. 307
Melsa J. L. 307
Michtchenko E. F. 299
Middleton D. 30, 300
Milenki A. V. 307
Mises R. 58, 230
Modestino J. W. 302
Moïsséev N. N. 299
Monro S. 294, 309

Nadaraïa E. J. 237, 293, 309
Nedoma N. 309
Noether G. E. 302
Nevelson M. B. 306, 309
Neyman J. 18*, 306

Okatomo M. 307

Parzen E. 237, 293
Patrick E. A. 307
Pearson K. 18*
Pinski A. I. 303, 306, 308
Pissarenko V. F. 306
Pitmann 300
Poisson S. D. 232*
Pontriaguine L. S. 299
Pougatchev V. S. 306
Prokofiev V. N. 308
Puri M. 302

Rao S. R. 301, 308, 309
Rayleigh D. W. S. 151*
Renyi A. 57, 302
Répine V. G. 30, 305, 306
Robbins H. 229, 294, 306, 309
Rosonoer L. I. 305
Roukhine A. L. 308
Roussas G. G. 304
Rubin H. 302
Rushforth C. K. 302
Rybine A. K. 303, 304

Sage A. P. 307
Sakrison D. L. 309
Samsonenko S. V. 308
Savage L. J. 299, 302
Scharf L. L. 309
Scuder H. J. 307
Sebastyen G. 306
Seidler J. 307
Sen P. K. 302, 304

Shinakov Yu. S. 307
Sidak Z. 301, 303
Skvortsov G. I. 308, 309
Smirnov N. V. 54, 301
Sossouline Yu. G. 30, 300, 306
Spearman 109
Spragins J. 307
Stein 287
Stuart A. 300, 301, 308, 309
Student 264

Tarassenko E. P. 301
Tartakovski G. P. 300, 305, 306
Tchervonenkis A. J. 305
Tchibissov D. M. 300, 304
Thomas J. B. 301, 302
Toussaint G. T. 307
Troïtski E. B. 305
Tsipkine Ia. Z. 299, 306

Van der Waerden B. L. 86 *, 301
Vapnik V. N. 305
Venter J. H. 309
Ventzel E. S. 299

Wagner T. J. 307
Wald A. 300
Wasan M. T. 305
Wiener N. 23 *
Wilcoxon 61 *
Wilks S. S. 304, 306
Williams T. R. 302
Wishart J. 211, 221
Witni 61
Woïnsky M. N. 302
Wolf S. S. 302
Wolfowitz J. 296, 309
Wolverton C. T. 307
Woodward P. W. 299

Zadeh 13, 299
Zagorouïko N. G. 305

# TABLE DES MATIÈRES

Préface à l'édition française . . . . . 5

Introduction . . . . . 7

**Chapitre premier. PROBLÈME DE L'ÉLIMINATION DE L'INDÉTERMINATION A PRIORI** 15

1.1. Problèmes de la synthèse avec information a priori complète 15
Terminologie (15). Principaux types de problèmes de la synthèse statistique (16). Critères de qualité et données a priori nécessaires (détection et discrimination des signaux) (17). Critères de qualité et données a priori indispensables (estimation des paramètres) (19). Types d'algorithmes (20). Fonction de décision (21). Modèles du signal et du bruit (22). Algorithmes optimaux de détection et de discrimination des signaux (25). Algorithmes optimaux de détection et de discrimination des signaux noyés dans un bruit normal additif (26). Estimation des paramètres et filtrage (29). Détection simultanée et estimation des paramètres des signaux (30). Résumé (31).

1.2. Problèmes de la synthèse dans les conditions de l'indétermination a priori . . . . . 31
Indétermination a priori paramétrique et non paramétrique (31). Algorithmes adaptatifs (33). Critères de qualité des algorithmes de détection des signaux dans les conditions d'indétermination paramétrique (33). Critères de qualité des algorithmes de détection des signaux dans les conditions d'indétermination non paramétrique (35). Critères de qualité des algorithmes adaptatifs (37). Critères de qualité asymptotiques (38). Algorithmes localement optimaux de détection des signaux (39). Algorithmes asymptotiquement optimaux de détection des signaux (41). Coefficient d'efficacité asymptotique relative (43). Algorithmes de détection des signaux asymptotiquement les plus efficaces (46). Calcul du coefficient d'efficacité asymptotique relative (47). Coefficient d'efficacité asymptotique relative des algorithmes unilatéraux utilisant les statistiques asymptotiquement normales (47). Coefficient d'efficacité asymptotique relative des algorithmes bilatéraux utilisant des statistiques asymptotiquement normales (50). Mesure quantitative de stabilité de l'algorithme asymptotiquement optimal de détection des signaux (50). Résumé (51).

**Chapitre 2. MÉTHODES NON PARAMÉTRIQUES DE SYNTHÈSE STATISTIQUE** 52

2.1. Critères de conformité . . . . . 52
Vérification de l'hypothèse sur la forme de la fonction de répartition (52). Critères de conformité (53). Critères de conformité de Kolmogorov et de Smirnov (54). Critère de conformité de Renyi (57). Critère de conformité de Mises (58). Critère de conformité du $\chi^2$ (59). Vérification de l'hypotèse d'appartenance de deux échantillons à une même distribution (60).

2.2. Statistiques utilisées dans les algorithmes non paramétriques de détection des signaux . . . . . 62
Statistiques de signes (62). Statistiques d'ordre (63). Statistiques de rang (64). Volume des calculs que nécessite la réalisation des algorithmes de rang (67).

2.3. Algorithmes de signe de détection des signaux noyés dans des bruits indépendants . . . . . 67
Détecteur de signe unilatéral du signal constant (67). Détecteur de signe bilatéral d'un signal constant (70). Efficacité relative du détecteur de signe unilatéral (71). Efficacité relative du détecteur de signe bilatéral (74). Détecteur de signe d'un signal déterministe quelconque (76). Efficacité relative du détecteur de signe d'un signal déterministe (78). Détecteur signe-quantiles d'un signal constant (79).

2.4. Algorithmes de rang de détection des signaux noyés dans des bruits indépendants . . . . . 80
Détecteur signe-rang d'un signal constant (80). Efficacité relative du détecteur signe-rang (84). Algorithmes signe-rang mixtes de détection d'un signal (87). Efficacité relative de l'algorithme signe-rang mixte (88). Détecteurs signe-rang d'un signal déterministe quelconque (90). Algorithmes de rang de détection des signaux à bande étroite (93). Coefficient d'efficacité asymptotique relative de l'algorithme de Wilcoxon (95). Algorithme de rang simplifié de détection des signaux (98).

2.5. Algorithmes non paramétriques de détection des signaux stochastiques noyés dans des bruits indépendants . . . . . 100
Système optimal à deux canaux de détection d'un signal normal noyé dans un bruit normal (100). Corrélateur de coïncidence des polarités (103). Efficacité relative du corrélateur de coïncidences des polarités (105). Corrélateur modifié de coïncidence des polarités (108). Algorithmes de rang de détection d'un signal stochastique (108).

2.6. Algorithme non paramétrique de détection d'un signal en présence d'un échantillon d'apprentissage du bruit . . . 110
Position du problème (110). Algorithme optimal de détection dans le cas d'un bruit normal additif (111). Algorithme de Wilcoxon à deux échantillons (112). Efficacité asymptotique relative de l'algorithme de Wilcoxon à deux échantillons (114).

2.7. Algorithmes de signe et de rang de détection des signaux noyés dans des bruits corrélés . . . . . 115
Modèle autorégressif des bruits corrélés (115). Algorithme de signe compte tenu de l'autorégression (116). Normalité asymptotique de la statistique (117). Efficacité relative de l'algorithme de signe de détection d'un signal pour un bruit

corrélé (118). Comparaison des efficacités des intégrations avec limitation et sans limitation (121). Algorithme signe-rang de détection d'un signal constant noyé dans un bruit corrélé (123). Algorithme de rang de détection d'un signal stochastique noyé dans un bruit corrélé (124).

2.8. Problèmes . . . . . 124

**Chapitre 3. ALGORITHMES ASYMPTOTIQUEMENT OPTIMAUX DE DÉTECTION DES SIGNAUX** 128

3.1. Structure des algorithmes asymptotiquement optimaux de détection des signaux d'après des observations indépendantes . . . . . 128
Développement asymptotique du logarithme du rapport de vraisemblance (128). Schéma de la démonstration du développement asymptotique (131). Algorithme asymptotiquement optimal de détection d'un signal déterministe (133). Caractéristique du convertisseur linéaire pour un bruit additif (135). Stabilité de l'algorithme asymptotiquement optimal de détection d'un signal déterministe (137). Algorithme asymptotiquement optimal de détection d'un signal quasi déterministe (141). Interprétation de l'algorithme asymptotiquement optimal de détection d'un signal quasi déterministe noyé dans un bruit indépendant (144). Algorithme asymptotiquement optimal de détection d'un signal modulé de phase aléatoire (145). Algorithme postdétecteur asymptotiquement optimal de détection d'un signal modulé de phase aléatoire (méthode d'amplitude) (149). Algorithme asymptotiquement optimal postdétecteur de détection d'un signal modulé de phase aléatoire (méthode de phase) (151). Distribution normale des paramètres du signal (155). Détection d'un signal stochastique (158).

3.2. Algorithmes de rang asymptotiquement optimaux de détection des signaux noyés dans des bruits indépendants . . . 160
Equivalence asymptotique des statistiques de rang et des statistiques non ordonnées (160). Algorithme de rang asymptotiquement optimal de détection d'un signal déterministe (163). Algorithmes de rang asymptotiquement optimaux de détection d'un signal déterministe noyé dans un bruit additif (165). Algorithme asymptotiquement optimal signe-rang de détection d'un signal déterministe (167). Stabilité de l'algorithme de rang asymptotiquement optimal de détection d'un signal déterministe (169). Comparaison des efficacités des algorithmes de rang et simples asymptotiquement optimaux de détection des signaux (170).

3.3. Algorithmes asymptotiquement optimaux de détection des signaux d'après des observations indépendantes quantifiées 171
Quantification en amplitude (171). Développement asymptotique du logarithme du rapport de vraisemblance pour des échantillons quantifiés (171). Algorithme asymptotiquement optimal de détection d'un signal déterministe (173). Détection d'un signal constant (174). Algorithme optimal simplifié de détection d'un signal déterministe d'après des observations quantifiées (176). Quantification optimale (176).

3.4. Structure des algorithmes asymptotiquement optimaux de détection des signaux d'après des observations corrélées . . . 178
Modèle de bruits corrélés (178). Développement asymptotique

du logarithme du rapport de vraisemblance (178). Algorithmes asymptotiquement optimaux de détection d'un signal déterministe noyé dans un bruit markovien $k$ fois connexe (181). Réalisation pratique de l'algorithme asymptotiquement optimal de détection d'un signal déterministe (182). Détection d'un signal déterministe noyé dans un bruit markovien (183). Algorithme asymptotiquement optimal de détection d'un signal quasi déterministe noyé dans un bruit markovien multiplement connexe (186). Réalisation de l'algorithme asymptotiquement optimal de détection d'un signal quasi déterministe (187). Stabilité de l'algorithme asymptotiquement optimal de détection d'un signal quasi déterministe noyé dans un bruit markovien multiplement connexe (188). Algorithme asymptotiquement optimal de détection d'un signal modulé de phase aléatoire (190). Distribution normale des paramètres d'un signal (192). Détection d'un signal stochastique (195). Algorithmes de rang asymptotiquement optimaux de détection des signaux noyés dans des bruits corrélés (196). Algorithme analogique asymptotiquement optimal de discrimination des signaux (197).

3.5. Problèmes . . . . . 199

Chapitre 4. ALGORITHMES ADAPTATIFS 205

4.1. Classification des observations dans le cas des distributions normales . . . . . 205
Algorithme de classification avec apprentissage (205). Cas unidimensionnel (206). Cas multidimensionnel (la matrice des covariances est donnée) (209). Cas multidimensionnel (la matrice des covariances est inconnue) (210). Cas multidimensionnel (les matrices des covariances sont différentes et inconnues) (211). Classification à alternatives multiples pour des distributions quelconques (212). Méthode géométrique (214). Algorithme de classification avec auto-apprentissage (215).

4.2. Propriétés asymptotiques des statistiques de classification lorsque le nombre d'indices informationnels augmente . . . 217
Position du problème (217). Cas de la symétrie sphérique des classes (218). Classification dans le cas d'indices corrélés (la matrice des covariances est donnée) (219). Classification dans le cas d'indices corrélés (la matrice des covariances est inconnue) (221).

4.3. Procédures d'apprentissage de Bayes . . . . . 222
Méthode des probabilités a posteriori (222). Détection d'un signal aléatoire noyé dans un bruit normal (cas unidimensionnel) (224). Détection d'un signal aléatoire corrélé noyé dans un bruit corrélé (226). Méthode euristique (227). Algorithme adaptatif bayesien de Robbins (229). Exemple d'algorithme de Robbins (231). Recherche du minimum du risque moyen à partir d'échantillons d'apprentissage (233). Algorithme récurrent de discrimination de deux signaux inconnus (234). Utilisation de la méthode des fonctions potentielles (236). Approximation de la statistique de classification par la méthode de Parzen-Nadaraïa (237). Méthode adaptative d'élimination de l'indétermination a priori des paramètres perturbateurs (238).

4.4. Algorithmes adaptatifs asymptotiquement optimaux de détection des signaux noyés dans des bruits . . . . . 240
Principes de construction des algorithmes adaptatifs asymptotiquement optimaux (240). Algorithme adaptatif asymptotiquement optimal de détection d'un signal déterministe noyé dans un bruit markovien $k$-fois connexe (241). Algorithme adaptatif asymptotiquement optimal de détection d'un signal quasi déterministe noyé dans un bruit markovien $k$-fois connexe (242). Algorithme adaptatif asymptotiquement optimal de détection d'un signal déterministe noyé dans un bruit additif à valeurs indépendantes (indétermination paramétrique) (243). Algorithme adaptatif asymptotiquement optimal de détection d'un signal déterministe noyé dans un bruis additif à valeurs indépendantes (indétermination non paramétrique) (243). Algorithme adaptatif asymptotiquement optimal de détection d'un signal déterministe noyé dans un bruit additif markovien (244). Détection d'un signal quasi déterministe noyé dans un bruit additif markovien multiplement connexe, satisfaisant à l'équation d'autorégression (245).
4.5. Propriétés asymptotiques des estimations de Bayes des paramètres du signal . . . . . 247
Position du problème (247). Estimation bayesienne de l'amplitude d'un signal noyé dans un bruit normal additif pour un temps d'observation grand (247). Propriétés asymptotiques de l'estimation bayesienne du paramètre vectoriel d'un signal noyé dans un bruit normal additif (249). Propriétés asymptotiques de la distribution a posteriori du paramètre vectoriel d'un signal (251). Propriétés asymptotiques des estimations de Bayes du paramètre vectoriel d'un signal pour des classes larges des bruits et des fonctions de pertes (253).
4.6. Problèmes . . . . . 254

**Chapitre 5. PRINCIPES DE SIMILITUDE ET D'INVARIANCE DANS LA SYNTHÈSE STATISTIQUE** 257

5.1. Principes de similitude et d'invariance . . . . . 257
Rôle des principes de similitude et d'invariance dans la théorie de la détection des signaux (257). Famille complète de densités de probabilité (258). Algorithmes semblables de Neyman (260). Algorithmes optimaux de Neyman (260). Règles invariantes de choix d'une décision lors de la vérification des hypothèses (262).
5.2. Détection d'un signal noyé dans un bruit normal additif de puissance inconnue . . . . . 263
Statistiques suffisantes dans le problème de la détection d'un signal déterministe (263). Règle sans biais uniformément la plus puissante de détection d'un signal déterministe (265). Détection d'un signal constant (265). Utilisation du principe d'invariance (267). Structure et caractéristique de fonctionnement du détecteur d'un signal déterministe (269). Détection d'un signal à bande étroite noyé dans un bruit normal additif de puissance inconnue (271). Règle invariante de détection d'un signal à bande étroite (273).
5.3. Détection d'un signal noyé dans un bruit normal additif de fonction de corrélation inconnue . . . . . 276

Détection d'un signal déterministe noyé dans un bruit quasi déterministe (276). Détermination du seuil (278). Règle invariante de détection (278). Détection d'un signal harmonique dans un bruit normal à paramètres inconnus du spectre énergétique (279).
5.4. Algorithmes de classification semblables et invariants . . . Position du problème (281). Algorithme optimal de classification de variables aléatoires normales unidimensionnelles (282). Algorithme de classification invariant uniformément le plus puissant (284). Classification des ensembles normaux multidimensionnels (285).
5.5. Problèmes . . . . . 287
Annexes . . . . . 290
1. Estimation d'une densité de probabilité inconnue . . . 290
2. Approximation stochastique . . . . . 294
3. Tables . . . . . 297

Bibliographie . . . . . 299
Notations utilisées . . . . . 310
Index alphabétique des matières . . . . . 313
Index des noms . . . . . .

## A NOS LECTEURS

Les Editions Mir vous seraient très reconnaissantes de bien vouloir leur communiquer votre opinion sur le contenu de ce livre, sa traduction et sa présentation, ainsi que toute autre suggestion.

Notre adresse: Editions Mir, 2, Pervi Rijski péréoulok, Moscou, I-110, GSP, U.R.S.S.

*Imprimé en Union Soviétique*